

Anna K. Naumova
Celia M.T. Greenwood *Editors*

Epigenetics and Complex Traits



Springer

Epigenetics and Complex Traits

Anna K. Naumova • Celia M.T. Greenwood
Editors

Epigenetics and Complex Traits

 Springer

Editors

Anna K. Naumova
Departments of Obstetrics
and Gynecology and Human Genetics
McGill University
Montreal, QC, Canada

Celia M.T. Greenwood
Lady Davis Institute for Medical Research
Centre for Clinical Epidemiology
Jewish General Hospital
Montreal, QC, Canada

ISBN 978-1-4614-8077-8

ISBN 978-1-4614-8078-5 (eBook)

DOI 10.1007/978-1-4614-8078-5

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013946543

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Influence of Epigenetic Phenomena on Gene Expression and Inheritance of Phenotypes

One of the many definitions of an epigenetic mark is a heritable feature that does not change the DNA sequence but determines when, where, and to what extent a gene will be expressed. Hence, epigenetics is a science that studies DNA packaging and regulation of its expression. Although often introduced as a new science, epigenetics dates back to the discovery of the roles of chromatin and DNA methylation in controlling gene expression in the 60s and 70s of the last century. Despite the intimate relationship between DNA and epigenetic factors, mainstream studies of genetic traits in humans and animal models have largely ignored the existence of epigenetic factors during the past decades, while the epigenetics community, although part of both the genetics and developmental biology fields, was digging deeper and deeper into the molecular mechanisms of epigenetic phenomena but seldom tackling problems of complex genetic traits in mammals. One of the reasons for the dichotomy is the very complexity of complex traits where small effects from multiple loci define the phenotype, whereas traditional molecular biology research required focusing on one selected target at a time. Another reason was the lack of methodologies capable of analyzing large amounts of epigenetic information in large cohorts of patients and controls. Nevertheless, during the last two decades, in-depth analysis of inheritance patterns combined with molecular approaches in a number of animal models, such as agouti viable yellow mice and callipyge sheep, has provided remarkable examples of how the interplay between genetic and epigenetic factors can generate complex traits.

Rapid technological improvements are now making it possible to measure epigenetic signals at many genomic locations in an unprecedented way and conduct prior-hypothesis-free epigenetic studies. Global initiatives such as the International Human Epigenome Consortium are underway to obtain high-resolution maps of histone modifications, DNA methylation, and transcription start sites and to compare epigenome signals and the resulting transcriptional regulation in a wide variety

of tissues and different cell types. However, even hypothesis-free data analyses require knowledge of epigenetic paradigms to make informed decisions when interpreting these massive data sets.

In this book, we have focused on the relationship between epigenetics and complex traits, since this field can be daunting for those wishing to do research. The biology is complex, and the ramifications of epigenetic regulation are widespread. Epigenetic states may contribute to the penetrance of genetic polymorphisms or mutations and thereby modify inheritance patterns. This may result in apparently non-Mendelian inheritance of genetic traits. Epigenetic changes in an individual may affect several different generations, depending on when these changes occur and in which cells. Genetic factors will influence epigenetic factors, and possibly their transmission. Effects may vary depending on sex, and also on the sex of an implicated parent. Concepts that applied in genetics, such as heritability, or the proportion of variance explained by genetics, can now be expanded to explicitly consider the epigenetic contributions. Furthermore, of course, different loci may demonstrate different associations with all these factors. Design of experiments and analysis of experimental data must reflect this complexity and be carefully approached.

Therefore, this book presents 14 detailed and distinct views on the interplay between complex traits and epigenetics. The chapters are grouped into three sections: (1) Fundamental aspects of the biology in epigenetics, with focus on the period in mammalian development that is pivotal for genetic transmission, i.e., gametogenesis and early embryonic development, insight into how the epigenetic marks are established, maintained, and transmitted and their influence on gene expression; (2) The known impact of epigenetic factors on several different complex traits and diseases of interest for human genetics; and (3) Approaches to experimental design and statistical analysis in this context.

Our hope is that the two communities of basic researchers and analysts will find mutual enrichment through this combination of material. An overview of available analytic methods and their underlying assumptions could inform experimental design choices. Similarly, improved understanding of the biology could lead to better choices for analysis, and an appreciation for the many factors that may need to be considered. Ultimately, this marriage of topics could lead to improved study designs, rich and complete analytic frameworks, new approaches to analysis, and guidelines for interpretation.

Of course, this book includes only a small overview of the available knowledge and approaches, yet we anticipate that this will be a helpful first reference for researchers entering the field, and will stimulate future developments. We thank Springer for making this endeavor possible.

Montreal, QC, Canada
Montreal, QC, Canada

Anna K. Naumova
Celia M.T. Greenwood

Acknowledgements

We sincerely thank Jason Apostolopoulos for all his assistance in putting this book together. Celia Greenwood is grateful to the Pharmaprix Weekend to End Women's Cancers (2010) for their support.

Contents

Part I Epigenetic Phenomena in the Germ Line and Early Embryonic Development and Their Effects on the Inheritance of Genetic Traits

- 1 Epigenetic Reprogramming in the Mammalian Germline** 3
Stéphanie Maupetit-Méhouas, David Nury, and Philippe Arnaud
- 2 Establishment of Tissue-Specific Epigenetic States During Development** 35
Ionel Sandovici
- 3 X-Chromosome Inactivation** 63
Wendy P. Robinson, Allison M. Cotton, Maria S. Peñaherrera, Samantha B. Peeters, and Carolyn J. Brown
- 4 *Cis*- and *Trans*-Effects Underlying Polar Overdominance at the Callipyge Locus** 89
Michel Georges, Haruko Takeda, Huijun Cheng, Xu Xuewen, Tracy Hadfield-Shay, Noelle Cockett, and Carole Charlier
- 5 Transgenerational Epigenetic Effects and Complex Inheritance Patterns** 107
Anna K. Naumova
- 6 Autosomal Monoallelic Expression** 131
Virginia Savova and Alexander A. Gimelbrant

Part II Epigenetic Variation in Health and Disease

- 7 Recurrent CNVs in the Etiology of Epigenetic Neurodevelopmental Disorders** 147
Janine M. LaSalle and Mohammad Saharul Islam

8	Impact of the Early-Life Environment on the Epigenome and Behavioral Development	179
	Benoît Labonté and Gustavo Turecki	
9	Interaction Between Genetics and Epigenetics in Cancer	209
	Amanda Ewart Toland	
 Part III Impact of Epigenetics on Complex Trait Genetics and Analysis		
10	Epigenetic Variation, Phenotypic Heritability, and Evolution	233
	Robert E. Furrow, Freddy B. Christiansen, and Marcus W. Feldman	
11	Statistical Approaches for Detecting Transgenerational Genetic Effects in Humans	247
	Janet S. Sinsheimer and Michelle M. Creek	
12	Transmission Ratio Distortion: A Neglected Phenomenon with Many Consequences in Genetic Analysis and Population Genetics	265
	Aurélie Labbe, Lam Opal Huang, and Claire Infante-Rivard	
13	Epigenome-Wide Association Studies: Potential Insights into Human Disease	287
	Christopher G. Bell	
14	Analytical Considerations for Epigenome-Wide Association Scans of Complex Traits	319
	Jordana T. Bell	
	Index	339

Contributors

Philippe Arnaud Genetic Reproduction and Development, Clermont Université-INSERM, Clermont-Ferrand, France

Christopher G. Bell Medical Genomics, UCL Cancer Institute, University College London, London, UK

Jordana T. Bell Department of Twin Research and Genetic Epidemiology, King's College, London, UK

Carolyn J. Brown Department of Medical Genetics, Molecular Epigenetics Group, University of British Columbia, Vancouver, BC, Canada

Carole Charlier Unit of Animal Genomics, University of Liège, Liège, Belgium

Huijun Cheng Unit of Animal Genomics, University of Liège, Liège, Belgium

Freddy B. Christiansen Department of Bioscience and Bioinformatics Research Center, University of Aarhus, Aarhus, Denmark

Noelle Cockett Department of Animal, Dairy, and Veterinary Sciences, Utah State University, Logan, UT, USA

Allison M. Cotton Department of Medical Genetics, Molecular Epigenetics Group, University of British Columbia, Vancouver, BC, Canada

Michelle M. Creek Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, CA, USA

Marcus W. Feldman Department of Biology, Stanford University, Stanford, CA, USA

Robert E. Furrow Department of Biology, Stanford University, Stanford, CA, USA

Michel Georges Unit of Animal Genomics, University of Liège, Liège, Belgium

Alexander A. Gimelbrant Department of Genetics, Harvard Medical School, Boston, MA, USA

Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA

Tracy Hadfield-Shay Department of Animal, Dairy, and Veterinary Sciences, Utah State University, Logan, UT, USA

Lam Opal Huang Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada

Claire Infante-Rivard Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada

Mohammad Saharul Islam Medical Microbiology and Immunology, Genome Center, MIND Institute, University of California, Davis, CA, USA

Aurélie Labbe Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada

Douglas Mental Health University Institute, Montréal, QC, Canada

Benoit Labonté McGill Group for Suicide Studies, Department of Psychiatry, Douglas Mental Health Institute, Montreal, QC, Canada

Janine M. LaSalle Medical Microbiology and Immunology, Genome Center, MIND Institute, University of California, Davis, Davis, CA, USA

Stéphanie Maupetit-Méhouas Genetic Reproduction and Development, Clermont Université-INSERM, Clermont-Ferrand, France

Anna K. Naumova Departments of Obstetrics and Gynecology and Human Genetics, McGill University, Montreal, QC, Canada

The Research Institute of the McGill University Health Centre, McGill University, Montreal, QC, Canada

David Nury Genetic Reproduction and Development, Clermont Université-INSERM, Clermont-Ferrand, France

Maria S. Peñaherrera Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Child & Family Research Institute, Vancouver, BC, Canada

Samantha B. Peeters Department of Medical Genetics, Molecular Epigenetics Group, University of British Columbia, Vancouver, BC, Canada

Wendy P. Robinson Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

Child & Family Research Institute, Vancouver, BC, Canada

Ionel Sandovici Metabolic Research Laboratories, Department of Obstetrics and Gynaecology, University of Cambridge, Cambridge, UK

Centre for Trophoblast Research, University of Cambridge, Cambridge, UK

Virginia Savova Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA

Department of Genetics, Harvard Medical School, Boston, MA, USA

Janet S. Sinsheimer Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

Haruko Takeda Unit of Animal Genomics, University of Liège, Liège, Belgium

Amanda Ewart Toland Department of Internal Medicine, Division of Human Genetics, Columbus, OH, USA

Gustavo Turecki McGill Group for Suicide Studies, Department of Psychiatry, Douglas Mental Health Institute, Montreal, QC, Canada

Xu Xuewen Unit of Animal Genomics, University of Liège, Liège, Belgium

Abbreviations

3C	Chromosome conformation capture
5C	Chromosome conformation capture carbon copy
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5hmU	5 Hydroxymethyluracil
5mC	5-Methylcytosine
aCGH	Array comparative genomic hybridization
ACTH	Adreno corticotropic hormone
ADHD	Attention deficit hyperactivity disorder
AdoMet	S-adenosyl-l-methionine
AEBP2	AE binding protein 2
AGRE	Autism Genetic Resource Exchange
AHEAD	Alliance for the Human Epigenome and Disease
AID or AICDA	Activation-induced cytidine deaminase
AMER1	APC membrane recruitment protein 1
AML	Acute myeloid leukemia
APOBEC1	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1
AR	Androgen receptor
ART	Assisted reproductive technologies
AS	Angelman syndrome
ASD	Autism spectrum disorder
ASE	Allele-Specific Expression
ASHM	Allele-Specific Histone Modifications
ASM	Allele-Specific Methylation
AVP	Vasopressin
Axin ^{Fu}	Axin fused allele
BDNF	Brain-Derived Neurotrophic Factor
BEGAIN	Brain-enriched guanylate kinase-associated protein

BER	Base excision repair
BiS-seq	Bisulfite 2nd generation sequencing
BLIMP1	B-lymphocyte induced maturation protein 1
BMI	Body Mass Index
BMI	Body mass composition
BMP	Bone morphogenetic protein
CBX5	Chromobox homolog 5
CCNE1	Cyclin E1
CDC25A	Cell division cycle 25A
CDX1	Caudal type homeobox 1
CDX2	Caudal type homeobox 2
CGI	CpG Island
CHD1	Chromodomain helicase DNA binding protein 1
ChIA-PET	Chromatin interaction analysis with paired-end tag sequencing
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
<i>CLPG</i>	Callipyge locus
CMCM	Case-mother, Control-mother
CNV	Copy number variation
CPA	Child physical abuse
CpG	Cytosine-phosphate-guanine
CPT	Case-parent trio
CRF	Corticotropin-releasing factor
CRH	Corticotropin-releasing hormone
CSA	Child sexual abuse
CTCF	CCCTC-binding factor [zinc finger protein]
D3	DIO3, type 3 deiodinase
DD	Developmental delay
DEX	Dexamethasone
DGS	DiGeorge syndrome
DHS	DNase I Hypersensitivity Sites
DKK1	Dickkopf 1 homolog (<i>Xenopus laevis</i>)
DLK1	Delta-like homologue 1; also known as preadipocyte factor-1 (PREF1) or fetal antigen (FA1)
DMD	Duchenne muscular dystrophy
DMR	Differentially Methylated Region
DNMT	DNA methyltransferase
DNMT1	DNA methyltransferase 1
DNMT1o	Oocyte-specific form of the DNA cytosine methyltransferase 1
DNMT3A	DNA methyltransferase 3A
DNMT3L	DNA methyltransferase 3-like protein
Dscam	Down Syndrome Cell Adhesion Molecule

DSL	Delta-Serrate-LAG-2 domain
DSL	Disease susceptibility locus
DZ	Dizygotic
EED	Embryonic ectoderm development
EFNB1	Ephrin-B1
EHMT2	Euchromatic histone-lysine N-methyltransferase 2
ELF5	E74-like factor 5 [ets domain transcription factor]
EMFG	Extended Maternal-Fetal Genotype
EMSA	Electrophoretic mobility shift assay (EMSA)
ENCODE project	Encyclopedia of DNA Elements project
EOMES	Eomesodermin
eQTL	Expression quantitative trait loci
ERV1	Class I endogenous retrovirus 1
ES	Embryonic stem cells
EWAS	Epigenome Wide Association Study
EZH2	Enhancer of zeste homolog 2 (<i>Drosophila</i>)
F XIII	Factor XIII
FGF	Fibroblast growth factor
FISH	Fluorescence in situ hybridization
FMR1	Fragile X mental retardation 1
FOXA	Forkhead box A
G6PD	Glucose-6-phosphate dehydrogenase
GAD1	Glutamate decarboxylase 1 [brain, 67 kDa]
GADD45a	Growth-arrest and DNA-damage-inducible protein 45 α
GATA	GATA binding protein
GATA4	GATA binding protein 4
GATA6	GATA binding protein 6
GC	Germ cell
gDMR	Germline differentially methylated region
GR	Glucocorticoid receptor
GSK3B	Glycogen synthase kinase 3 beta
<i>GTL2</i>	Gene trap locus 2
GWAS	Genome-wide association study
HDAC	Histone deacetylase
HDN	Hemolytic disease of the newborn
HIRA	Histone cell cycle regulation defective homolog A (<i>S. cerevisiae</i>)
HLA-DQB1	Major histocompatibility complex, class II, DQ beta 1
HLA-DRB1	Major histocompatibility complex, class II, DR beta 1
HNF4A	Hepatocyte nuclear factor 4, alpha
HOXD11	Homeobox D11
HOXD12	Homeobox D12

HPA	Hypothalamic–pituitary–adrenal axis
HSM	Haplotype-Specific Methylation
HUMARA	Human androgen receptor
IAP	Intracisternal A particle
IBD	Identically by descent
ICM	Inner cell mass
ICR	Imprinting control region
IG DMR	Intergenic differentially methylated region
iPS	Induced pluripotent stem cells
iQTL	Imprinted QTL
JARID2	Jumonji, AT rich interactive domain 2—a member of the Jumonji family of lysine demethylases)
JMJD3/KDM6B	Lysine [K]-specific demethylase 6B
KDM1B	Lysine (K)-specific demethylase 1B
KDM5C	Lysine (K)-specific demethylase 5C
KLF2	Kruppel-like factor 2
KRAB	Kruppel-associated box
L1	LINE element, long interspersed repetitive element 1
LCR	Low copy repeat
LD	Linkage disequilibrium
LG	Licking and grooming
LIF/STAT3	Leukemia inhibitory factor/signal transducer and activator of transcription 3
lincRNAs	Large intergenic noncoding RNAs
LINE1	Long interspersed repeat element 1
<i>LIS1</i>	Lissencephaly-1 gene
LMR	Low Methylation Region
lncRNA	Long non-coding RNA
lncRNA	Long noncoding RNA
LOI	Loss of imprinting
LRT	Long-range transgenerational
LRT-M	Long range transgenerational effects on the maternal side
LRT-P	Long range transgenerational effects on the paternal side
LSH	Lymphoid-specific helicase
LTR	Long terminal repeat
MAE	Monoallelic expression
MAOA	Monoamine oxidase A
Mat-gDMR	Maternal-germline differentially methylated region
MBD3	Methyl-CpG binding domain protein 3
MBD4	Methyl CpG binding domain protein 4
MBD-seq	Methylated DNA binding domain sequencing
MDLS	Miller-Dieker syndrome
MDR	Methylation Determining Region

MECAP-seq	Methylated DNA capture by affinity purification sequencing
MECP2	Methyl CpG binding protein 2
Me-DIP	Methylated DNA immunoprecipitation
MeDIP-seq	Methylation Dependent Immunoprecipitation 2nd generation sequencing
MEK	MAP kinase/ERK kinase]
me-QTLs	Methylation quantitative trait loci
methOR	Methylation odds ratios
MFG	Maternal fetal genotype
MGMT	O ⁶ -methylguanine-DNA methyltransferase
MHC	Major histocompatibility region
Mirg	Micro-RNA containing gene (cluster of ~50 miRNAs expressed from the maternal allele)
miRNA	Micro RNA
MLL/Trithorax complex	Myeloid/lymphoid or mixed-lineage leukemia [trithorax homolog, Drosophila]
MLL3/KMT2C	Lysine (K)-specific methyltransferase 2C
MLL4/KMT2D	Lysine (K)-specific methyltransferase 2D
MNase	Micrococcal nuclease
M-PCR	Methylation-specific PCR
MRFs	Myogenic regulatory bHLH-containing factors
MSUC	Meiotic silencing of unsynapsed chromatin
MT	Mouse transcript
MTHFR	Methylenetetrahydrofolate reductase (NAD(P)H)
MVH	Mouse vasa homolog
MVP	Methylation Variable Position
MZ	Monozygotic twins
NANOG	Nanog homeobox
NAP-1	Nucleosome assembly protein-1
ncRNA	Noncoding RNA
ND	Neurodevelopmental disorders
NER	Nucleotide excision repair
Nes	Nestin
NF1	Neurofibromatosis type 1
NGF	Nerve growth factor
NIMA	Non-inherited maternal antigen
NIPA	Non-inherited paternal antigen
NIPBL	Nipped-B homolog [Drosophila]
NP95 (or UHRF1)	Nuclear protein of 95 kDa (or ubiquitin-like with PHD and ring finger domains 1)
NT3/4	Neurotrophin 3 and 4
NuRD	Nucleosome-remodeling

OATL1	Ornithine aminotransferase-like 1
OCT4	Octamer-binding transcription factor 4, also known as POU5F1 – POU domain, class 5, transcription factor 1
<i>Om</i>	Ovum mutant
PAI-1	Plasminogen activator inhibitor-1
PAR	Pseudoautosomal region
pasRNA	Promoter-associated small RNA
PAT	Parental Asymmetry Test
Pat-gDMR	Paternal-germline differentially methylated region
PcG	Polycomb group
PCL2	Polycomb-like 2 protein
PCR	Polymerase chain reaction
PCSK1N	Proprotein convertase subtilisin/kexin type 1 inhibitor
PFG	Paternal fetal genotype
PGC	Primordial germ cell
PGK1	Phosphoglycerate kinase 1
PGL/PCC	Paraganglioma/pheochromocytoma
PHF6	PHD finger protein 6
piRNA	PIWI-interacting RNA
POE	Parent of origin effects
POF	Premature ovarian failure
PO-LRT	Parent-of-Origin Likelihood Ratio Test
POMC	Pro opiomelanocortin
PoO	Parent of origin
PPB	Pleuropulmonary blastoma
PRC1	Repressive complex 1
PRC2	Polycomb repressive complex 2
PRDM	PR-domain-zinc-finger protein
PRMT5	Protein arginine N-methyltransferase 5
PTSD	Post traumatic stress disorder
PVN	Paraventricular nucleus
PWS	Prader-Willi syndrome
qTCAs	Transcriptional clonality assays
QTL	Quantitative trait locus
RASGRF1	RAS protein-specific guanine nucleotide-releasing factor 1
RFLP	Restriction fragment length polymorphism
RM	Recurrent miscarriage
RNA pol II	RNA polymerase II
RNAP II	RNA polymerase II
RNA-seq	RNA-sequencing
RNF2	Ring finger protein 2, also known as RING1B
RRBS	Reduced representation bisulfite sequencing
RRBS-seq	Reduced Representation BiSulfite 2nd generation sequencing

rRNA	Ribosomal RNA
RTL1	Retrotransposon-like 1
RTT	Rett syndrome
SALL4	Sal-like protein 4
SAM	S-adenosylmethionine
SAT1	Spermidine/spermine N1-acetyltransferase
Satb2	SATB homeobox 2
SETDB1	SET domain, bifurcated 1
SGA	Small-for-gestational-age
SINE	Short interspersed repeat element
siRNA	Small interfering RNA
SKI	v-ski sarcoma viral oncogene homolog (SKI)
Smarca5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5
SMS	Smith-Magenis syndrome
SMS	Spermine synthase
SMUG1	Single-strand-selective monofunctional uracil DNA glycosylase 1
sno-RNA	Small-nucleolar RNA
SNP	Single nucleotide polymorphism
SOX17	SRY [sex determining region Y]-box 17
Sox2	SRY-box containing gene 2
SOX2	SRY [sex determining region Y]-box 2
SOX7	SRY [sex determining region Y]-box 7
SRY	Sex-determining region Y
Ssm 1	Strain-specific modifier of transgene methylation 1
STAG2	Stromal antigen 2
STR	Short tandem repeat
STS	<i>STS</i> (steroid sulfatase (microsomal), <i>isozyme S</i>)
SUZ12	Suppressor of zeste 12 homolog [<i>Drosophila</i>]
T2D	Type 2 diabetes
TAT	Transmission Asymmetry Test
TCF7L2	Transcription factor 7-like 2 [T-cell specific, HMG-box]
TDG	Thymine DNA glycosylase
TDRD	Tudor domain
TDT	Transmission disequilibrium test
TE	Trophectoderm
TEAD4	TEA domain family member 4
TET	Ten-eleven-translocation
TET3	Ten-eleven translocation
<i>Tex19.1</i>	Testis expressed gene 19.1
TFBS	Transcription Factor Binding Sites
TGF	Transforming growth factor

TIMP1	TIMP metalloproteinase inhibitor 1
TIP60/ KAT5–P400	Lysine acetyltransferase 5/ E1A binding protein p400
tiRNA	Transcription initiation RNA
TRD	Transmission ratio distortion
TRIM28	Tripartite motif containing 28
TrkB	Tropomyosin-Related Kinase B
TS	Trophoblast stem cells
<i>TSG</i>	Tumor suppressor gene
TSSs	Transcription start sites
<i>UBE3A-AS</i>	Antisense transcript of <i>UBE3A</i>
UBF	Upstream binding factor
UCE	Upstream control element
UPD	Uniparental disomy
UTX/KDM6A	Lysine [K]-specific demethylase 6A
VCFS	Velo-cardiofacial syndrome
WBS	Williams-Beuren syndrome
WGAS	Whole Genome sequencing Association Study
WGBS	Whole genome bisulfite sequencing
XCI	X-chromosome inactivation
XEN	Extraembryonic endoderm stem cells
<i>XIST</i>	X-inactive specific transcript
<i>XIST/Xist</i>	Inactive X specific transcripts
ZDHHC15	DHHC-type containing 15
ZFP57	Zinc finger protein 57
ZNF274	Zinc finger protein 274

Part I
Epigenetic Phenomena in the Germ Line
and Early Embryonic Development and
Their Effects on the Inheritance
of Genetic Traits

Chapter 1

Epigenetic Reprogramming in the Mammalian Germline

Stéphanie Maupetit-Méhouas, David Nury, and Philippe Arnaud

Abstract Epigenetic modifications are crucial for maintaining and faithfully transmitting the identity of each cell type during cell division. During mammalian germ cell development, the acquisition of the ability to form a totipotent zygote is associated with extensive epigenetic reprogramming that affects all major developmental processes, including genomic imprinting, X-inactivation, retroelement silencing and gene expression. The existing epigenetic patterns are first erased during primordial germ cell development, followed by acquisition of a germline-specific epigenetic signature that can be eventually transmitted to and interpreted by the progeny. A better characterisation of the underlying mechanisms is relevant for both fundamental and clinical research dealing with epigenetic inheritance, epigenetic control of mammalian development and regenerative medicine. In this review we present and discuss recent advances on the nature, mechanisms and consequences of resetting the epigenetic pattern during primordial germ cell formation and (re)acquiring a new set of epigenetic marks at later stages of germline development.

1.1 Introduction

During somatic development of higher organisms, pluripotent cells progressively reduce their differentiation potential and become committed to a particular cell fate with specific gene expression and functional profiles. This tightly regulated process requires the concerted action of specific factors and is accompanied or caused by dynamic chromatin changes that influence gene expression patterns and phenotype.

S. Maupetit-Méhouas • D. Nury • P. Arnaud (✉)
GReD (Genetic Reproduction and Development), CNRS UMR6293-Clermont
Université-INSERM U1103, 28 Place Henri Dunant BP38,
63001 Clermont-Ferrand Cedex, France
e-mail: stephanie.maupetit_mehouas@udamail.fr; david.nury@u-clermont1.fr;
philippe.arnaud@udamail.fr

These changes occur at the level of DNA methylation, histone tail modifications, nucleosome remodelling and regulation of higher order chromatin structures. Most (but not all) of these modifications are heritable from one cell generation to the next and are thus referred as being epigenetic. Thus, each cell type in an organism is characterised by a specific and stable epigenetic signature (epigenome) that is transmitted to the daughter cells. Once specified, the epigenome of a cell type is relatively stable. However, in mammals, there are two key developmental stages in which epigenetic patterns are profoundly modified, with erasure of the existing epigenetic marks and acquisition of a new set. This so-called epigenetic reprogramming occurs first in early embryogenesis, following fertilisation, when the epigenetic information carried by the mature gametes is removed and replaced by an embryonic/somatic signature at the peri-implantation stage. This “embryonic” reprogramming is incomplete as some genomic regions, notably the *cis*-acting regulatory sequences of imprinted gene loci (imprinting control regions, ICRs), escape this process. A more thorough epigenetic reprogramming occurs during gametogenesis and it virtually impacts all epigenetic-based developmental processes: genomic imprinting, X-inactivation, retroelement silencing and gene expression. The understanding of the underlying mechanisms is relevant for both fundamental and clinical research. It will enable to better define the role of epigenetics in the control of mammalian development and also to elucidate the mechanism of in vitro-induced reprogramming/pluripotency.

This review focuses on the germline epigenetic reprogramming and discusses recent findings on the mechanisms involved in erasing the epigenetic pattern during primordial germ cell (PGC) formation and in (re)acquiring a new set of epigenetic marks at later stages of germline development.

1.2 Temporal and Spatial Dynamics of Mouse Germ Cell Development

Among all the cell lineages of a complex organism, only germ cells can give rise to a new individual, allowing the transmission of genetic and possibly epigenetic information to the next generation. Germ cell development initiates with the specification of PGCs, which following colonisation of the embryonic gonads will develop into oocytes or spermatozoa. In mammals, most of our knowledge on the temporal and spatial dynamics of this tightly regulated process comes from the mouse model (Fig. 1.1).

Unlike other non-mammalian species, such as *D. melanogaster* and zebrafish, mouse PGCs are not predetermined at fertilisation but are specified in the post-implantation embryo. At embryonic day 4.5 (E4.5), following blastocyst implantation, there is a rapid increase in the number of inner cell mass cells, leading to the formation of the epiblast (the source of all the body cell lineages). Germ cell fate

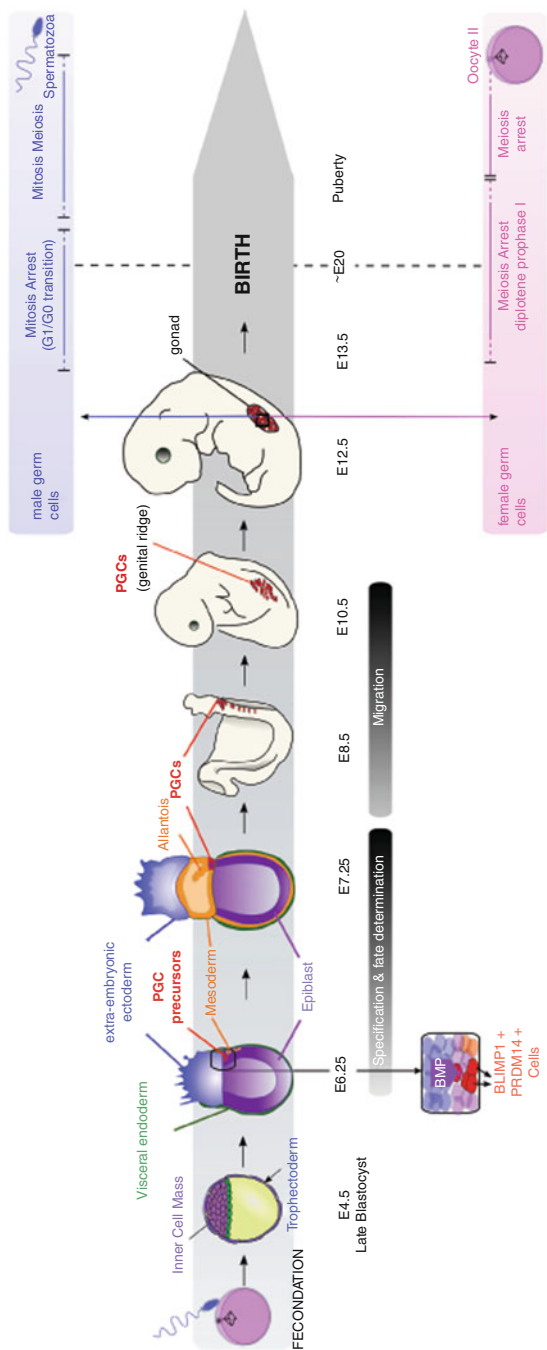


Fig. 1.1 Temporal and spatial dynamics of mouse germ cell development. Germ cell fate is induced at E6.25, and the lineage is determined around E7.25. From E7.5, primordial germ cells (PGCs) initiate migrating and then colonise the nascent genital ridge at around E10.5. At E12.5, PGCs start differentiating into male or female gametes. E4.5: Embryonic day 4.5. BMP: Bone morphogenetic protein. BLIMP1: B-lymphocyte maturation-induced protein 1. PRDM: PR domain zinc finger protein

is induced in the proximal epiblast in a dose-dependent manner by bone morphogenetic protein (BMP) signals from the extra-embryonic ectoderm at ~E6.25 (Lawson et al. 1999). This leads to the formation of a pool of PGC precursors of which only a limited number (about 6 cells), characterised by the expression of the zinc finger transcriptional regulators BLIMP1 (B-lymphocyte-induced maturation protein 1, also known as PR-domain-zinc-finger protein 1, PRDM1) and PRDM14 (PR-domain-zinc-finger protein 14), acquire a PGC fate. As a result, “fate-determined” PGCs emerge at ~E7.25 as a cluster of ~20–40 cells located at the base of the forming allantois (Ginsburg et al. 1990; Ohinata et al. 2005, 2009; Yamaji et al. 2008). From ~E7.5, PGCs migrate through the hindgut and mesentery and start colonising the nascent genital ridges (i.e., the future gonads) at ~E10.5. During this process, PGCs rapidly proliferate: from around 100 PGCs at E8.5 to ~200 at E9.5 and ~600 at E10.5. In the genital ridges, PGCs still proliferate up to E13.5 (~26 000 cells), when they stop dividing (Mochizuki and Matsui 2010; Kagiwada et al. 2012) (Fig. 1.2).

Following colonisation of the developing gonad, at E12.5, PGCs, now referred to as germ cells (GCs), start differentiating into male or female gametes. In the developing ovary, at E13.5, female GCs initiate meiosis I that will be blocked at the diplotene stage of prophase I at about the time of birth and until puberty. Following ovulation, the oocyte resumes meiosis I and halts in metaphase of meiosis II that will be completed after fertilisation (Smallwood and Kelsey 2012).

Conversely, male GCs do not initiate meiosis in the embryo and stop dividing from E13.5 (Western et al. 2008). At sexual maturity, male GCs will differentiate into spermatogonial stem cells and resume mitotic proliferation to form spermatocytes that will give rise, following meiosis, to haploid spermatids that will develop into spermatozoa.

1.3 Primordial Germ Cell Development and Reprogramming

After implantation, epiblast cells mature and prepare for gastrulation and formation of all the body cell lineages. This process is associated with major epigenetic changes, as illustrated by the genome-wide increase in DNA methylation in pre-gastrulating embryos that will be almost complete by E6.5 (Borgel et al. 2010). Thus, by ~E6.25, the prospective PGCs have accumulated several layers of epigenetic information and are already primed towards a somatic fate. Upon PGC specification, these epigenetic features will be erased through a major transcriptional and epigenetic reprogramming that might be important for the production of a totipotent zygote following fertilisation.

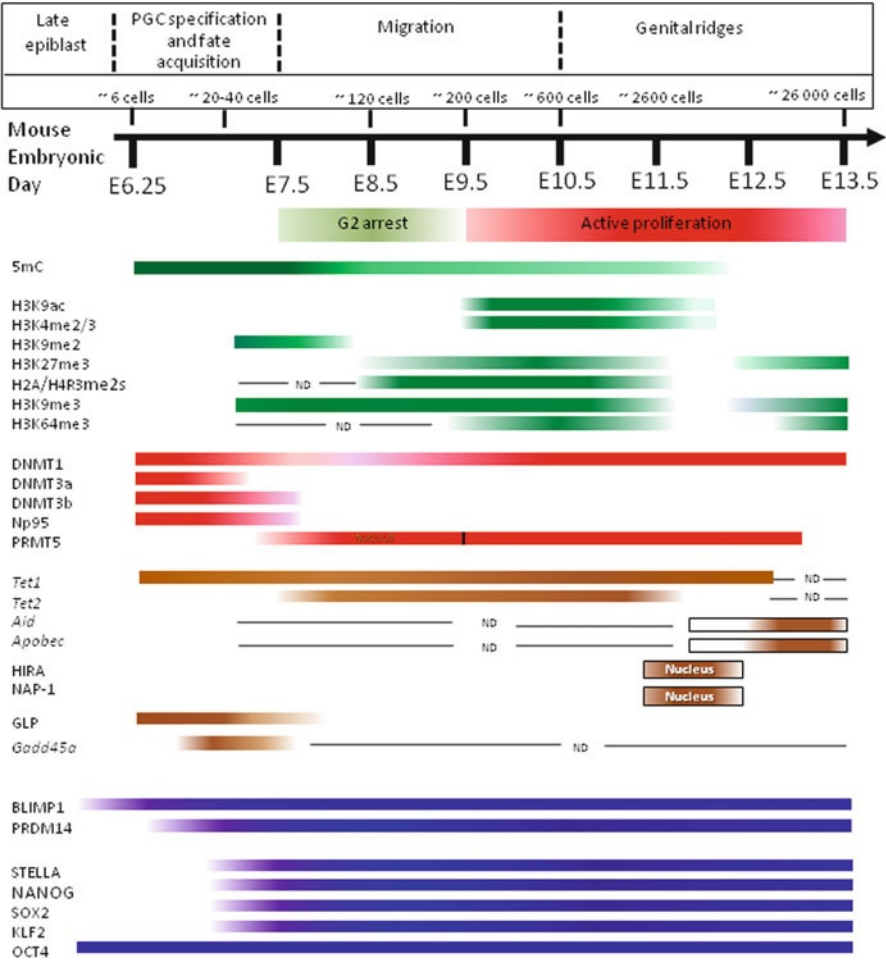


Fig. 1.2 Temporal schematic of epigenetic reprogramming during mouse primordial germ cell development. Genome-wide dynamics of DNA methylation and main histone modifications during PGC development (mainly revealed by immunochemistry analysis) are depicted. The dynamic expression of key epigenetic modifiers and pluripotency factors is also shown. Based on Kurimoto et al. (2008a, b); Ancelin et al. (2006); Seki et al. (2005, 2007); Hajkova et al. (2008, 2010); Daujat et al. (2009); and Hackett et al. (2013). PGC: Primordial germ cells, BLIMP1: B-lymphocyte-induced maturation protein 1, PRDM14: PR-domain-zinc-finger protein 14, SOX2: SRY (sex-determining region Y)-box 2, KLF2: Kruppel-like factor 2, OCT4 or POU5F1: POU class 5 homeobox 1, Dnmt: DNA (cytosine-5)-methyltransferase, NP95 or UHRF1: ubiquitin-like with PHD and ring finger domains 1, PRMT5: protein arginine methyltransferase 5, *Tet*: ten-eleven-translocation, *Aid* or *Aicda*: activation-induced cytidine deaminase, *Apobec*: apolipoprotein B mRNA editing enzyme catalytic polypeptide, HIRA: histone cell cycle regulation defective homolog A, NAP-1: nucleosome assembly protein 1, *Gadd45a*: growth arrest and DNA-damage-inducible protein 45 alpha

1.3.1 *Primordial Germ Cell Specification: Reprogramming Their Transcription Pattern*

PGC specification is associated with major changes in gene transcription to repress the somatic cell program and activate the germ cell-specific program, reacquire their pluripotency potential and prepare for the imminent genome-wide epigenetic reprogramming. This highly ordered process is regulated by BLIMP1 and PRDM14. At ~E6.25 these transcriptional regulators co-mark epiblast cells that will form PGCs and in the absence of either of these proteins, nascent PGC precursors fail to properly develop (Ohinata et al. 2005; Yamaji et al. 2008; Vincent et al. 2005; Kurimoto et al. 2008a, b). A single-cell microarray approach to establish the genome-wide transcription dynamics of developing PGCs and their somatic neighbours from E6.25 to E8.25 revealed that germ cell specification involves the up-regulation of nearly 500 “germ cell-specification” genes and the down-regulation of 330 “somatic program” genes (Kurimoto et al. 2008a). Among the down-regulated “somatic” genes there are many genes involved in embryonic development (e.g., *Hox* genes, *Dkk1*, *Cdx1* . . .), cell cycle regulation (e.g., *Ccne1*, *Cdc25a* . . .) as well as DNA methylation and histone modification, such as the de novo DNA methyltransferases DNMT3A and DNMT3B, the nuclear protein of 95 kDa (NP95, a factor essential to maintain the DNA methylation pattern during cell division) and the H3K9me2 histone methyltransferase GLP (G9a-like protein). Conversely, the “germ cell specification” category includes genes associated with germ cell development, such as *Stella* or *Fragilis*, and also the pluripotency genes *Nanog*, *Sox2* (*Sry-box2*) and *Klf2* (*Kruppel-like factor 2*) (Kurimoto et al. 2008a).

Further analysis conducted using BLIMP1-deficient PGC-like cells showed that BLIMP1 functions as a dominant repressor of the somatic program and is also involved in the reacquisition of the pluripotency potential and in the forthcoming epigenetic reprogramming. On the other hand, PRDM14 is required for *Sox2* up-regulation and *Glp* repression and is essential for the reacquisition of the pluripotency potential and for epigenetic reprogramming (Yamaji et al. 2008; Kurimoto et al. 2008b). Importantly, BLIMP1, although unnecessary to induce *Prdm14* expression, is strictly required for its maintenance (Yamaji et al. 2008).

How precisely these two proteins regulate germ cell specification remains to be established. Both BLIMP1 and PRDM14 contain a zinc-finger and histone methyltransferase SET domains, but no associated histone-modifying activity has been reported. Alternatively, they could exert their functions by recruiting effector partners to their target genes. BLIMP1 can recruit different chromatin-modifying proteins, such as histone deacetylases (HDAC) (Yu et al. 2000), G9A (Gyory et al. 2004) and the arginine methyltransferase PRMT5 (Ancelin et al. 2006). BLIMP1 and PRMT5 co-localise in the nuclei of migrating PGCs (Ancelin et al. 2006); however, it is not known whether the putative BLIMP1/PRMT5 complex is formed also during PGC specification and whether it contributes to repression of the somatic program.

Upon PGC specification, in addition to repressing the DNA methylation machinery and *Glp*, PGCs express factors that are currently used for in vitro somatic cell reprogramming. These include *Sox2*, *Nanog*, *Lin28* and *Klf2*, which are specifically up-regulated in PGCs, as well as *Oct-4* (*octamer-binding transcription factor 4*) which is strongly expressed already in epiblast cells and throughout PGC development (Kurimoto et al. 2008a; West et al. 2009; Yeom et al. 1996) (Fig. 1.2). The expression of these pluripotency factors might explain how monopotent PGCs can develop into pluripotent embryonic germ cells in culture (Matsui et al. 1992). In addition, it suggests that germ cell reprogramming shares some similarities with the mechanism underlying the generation of induced pluripotent stem cells (iPS). Nonetheless, the role of these pluripotency factors in PGC specification and particularly their potential reprogramming function remain to be determined (Gillich and Hayashi 2011). To date, germline-specific knockout experiments revealed that OCT-4 and NANOG are critical for PGC survival during migration (Kehler et al. 2004; Chambers et al. 2007; Yamaguchi et al. 2009).

1.3.2 Epigenetic Reprogramming in Early and Late Primordial Germ Cells

When the germ cell fate is established at ~E7.25, PGCs display an epigenetic pattern similar to that of the surrounding somatic cells (Seki et al. 2005, 2007; Hajkova et al. 2008; Popp et al. 2010; Guibert et al. 2012). Epigenetic reprogramming initiates at ~E7.5 and will lead to a virtually complete loss of DNA methylation in the germline of both sexes by E13.5 (Fig. 1.2).

Migrating PGCs undergo ordered epigenetic changes, leading to the establishment of a germ cell-specific chromatin signature which is distinct from that of the somatic neighbours. Immunocytochemistry approaches revealed that in PGCs most H3K9me₂, a repressive mark, is progressively removed in a cell-by-cell manner between E7.75 and E8.75 (Seki et al. 2005, 2007; Hajkova et al. 2008), while DNA methylation declines genome-wide (Seki et al. 2005, 2007). A recent study based on a genome-wide bisulphite sequencing approach further supports that the bulk of methylation erasure occurs prior to E9.5, showing an average methylation level at CpG sites of ~71 % in E6.5 epiblast cells and ~30 % in E9.5 PGCs (Seisenberger et al. 2012). Interestingly, some specific sequence classes, including X-linked, imprinted and some germline-specific genes, partially escape this initial loss of methylation (Seisenberger et al. 2012).

Conversely, the level of H3K9me₃, which marks centromeric heterochromatin, remains unchanged during PGC migration. H3K27me₃, another repressive mark that is mediated by the polycomb repressive complex 2 (PRC2), is progressively up-regulated from ~E8.25, with most PGCs showing high H3K27me₃ levels by E9.5. *Ezh2*, which encodes a histone methyltransferase belonging to PRC2, is concomitantly expressed in PGCs and could be involved in this up-regulation.

However, *Ezh2* is expressed also in somatic cells (Kurimoto et al. 2008a), suggesting that the removal of H3K9me2 and/or reduction of DNA methylation is a prerequisite for H3K27me3 acquisition. This hypothesis is supported by the observation that *Prdm14*^{-/-} PGCs, which cannot repress *Glp* and thus reduce H3K9me2, also fail to up-regulate H3K27me3 (Yamaji et al. 2008). Consequently, H3K27me3 enrichment should compensate for H3K9me2 reduction and ensure that the PGC euchromatin is properly repressed (Seki et al. 2005). The PRMT5-mediated symmetrical arginine methylation at histone H2A and H4 (H2A/H4R3me2), a repressive mark that increases from E8.5 onward through the action of the putative PRDM1–PRMT5 complex (Ancelin et al. 2006), might have a similar role. These dynamic histone modifications highlight that the euchromatic part of the PGC genome is deficient in repressive marks between ~E7.5 and ~E8.25. However, this “repression-free” chromatin state probably does not result in deregulated transcription as PGCs exhibit a transient loss of RNA polymerase II (RNA polII) activity during this period (Seki et al. 2007). Accordingly, enrichment for histone marks associated with transcriptionally active/permissive chromatin, such as H3K4me2/3 and H3K9ac, is observed only in late-migrating PGCs and further increases when they enter the genital ridges (Seki et al. 2005; Hajkova et al. 2008).

The down-regulation of the GLP histone methyltransferase and the DNA methylation machinery, which occurs during PGC specification, probably accounts for the dramatic decrease in H3K9me2 and DNA methylation observed by E8.75, through a passive process. However, the finding that most of the migrating PGCs stop dividing at the G2 phase of the cell cycle (G2 arrest) between E7.75 and E8.75 indicates that the loss of these repressive marks might also occur independently of DNA replication (see Sect. 1.4).

These dynamic and ordered epigenetic changes in migrating PGCs are believed to prepare their genome for the second wave of reprogramming events following their entry in the genital ridge at E10.5. This major reprogramming wave is first characterised by a second phase of demethylation (Seki et al. 2005; Hajkova et al. 2008) that affects the sequences that have retained their methylation in migrating PGCs (Seisenberger et al. 2012), leading to global genome hypomethylation in the germline of both sexes by E13.5. Genome-wide and region-specific bisulphite-based approaches indeed confirmed the removal of DNA methylation at the majority of the genome including coding and intergenic regions, X-linked, imprinted and germline-specific genes as well as most of transposable elements and centromeric regions (Hajkova et al. 2002; Maatouk et al. 2006; Yamagata et al. 2007; Popp et al. 2010; Henckel et al. 2011; Guibert et al. 2012; Seisenberger et al. 2012), revealing that methylation erasure in PGCs is more complete than in preimplantation embryos (Borgel et al. 2010; Smallwood et al. 2011). Moreover, the onset of DNA demethylation precedes the massive chromatin remodelling at ~E11.5 associated with transient loss of the linker histone H1 and loss of H3K27me3, H3K64me3 and H3K9me3 histone modifications that are subsequently regained by E12.5, while others, such as H3K9ac and H2a/H4R3me2s, are stably removed (Hajkova et al. 2008; Daujat et al. 2009).

The simultaneous disappearance of several histone modifications suggests that chromatin remodelling relies on a histone replacement process. Accordingly, HIRA and NAP-1, two chaperones involved in histone replacement, concomitantly accumulate in PGC nuclei (Hajkova et al. 2008). Histone demethylases could also be involved in this process. Indeed, in PGCs deficient for UTX (ubiquitously transcribed tetratricopeptide repeat gene on X chromosome), an H3K27me3 demethylase, H3K27me3 does not transiently decrease and these cells cannot develop properly (Mansour et al. 2012).

Thus, at around E12.5, following the removal of DNA methylation and several histone modifications, GCs display a basal and probably naïve epigenomic signature that is unique during mammalian development.

1.3.3 Epigenetic-Based Developmental Processes and Primordial Germline Reprogramming

Reprogramming erases the epigenetic program acquired by ~E6.25 in the epiblast-derived PGC precursors. Genomic imprinting resetting and X-chromosome reactivation in female GCs are the two main hallmarks of this phenomenon, which affects also gene expression and retroelement regulation. However, like for DNA methylation erasure that occurs in a stepwise manner, these events do not happen all at the same time.

In the female germline, X-chromosome reactivation was initially believed to occur upon colonisation of the genital ridges (Tam et al. 1994). More recent studies revealed that this event initiates already in nascent/migrating PGCs. It is triggered by progressive down-regulation of *Xist*, a noncoding RNA essential for X inactivation, from E7.0 to E10.5 (Sugimoto and Abe 2007) that leads to gradual loss of H3K27me3 enrichment on the inactive X chromosome (De Napoles et al. 2007). Reactivation of X-linked genes is completed in post-migratory GCs by E14.5 (Tam et al. 1994; Sugimoto and Abe 2007).

Genomic imprinting resetting is restricted to post-migratory GCs. ICRs, the *cis*-acting regions that control imprinted domains, retain most, though not all, of their DNA methylation marks up to E10.5 before undergoing rapid demethylation that is complete by E13.5 (Hajkova et al. 2002; Lee et al. 2002; Seisenberger et al. 2012; Kagiwada et al. 2012). Elegant studies conducted on cloned embryos produced from single PGC nuclei demonstrated that during this time window functional imprints are lost at different imprinted loci at different times. This timing correlates with erasure of DNA methylation at the associated ICRs (Kato et al. 1999; Lee et al. 2002).

Besides the effects on X-linked and imprinted genes, it is less clear whether PGC reprogramming and the associated extensive erasure of DNA methylation affect gene expression in general. A recent study based on an Me-DIP approach coupled to microarray analysis showed that only a limited number of promoters are methylated in E7.5 epiblast cells (Guibert et al. 2012), similarly to what is observed

in somatic cells (Weber and Schübeler 2007; Mohn et al. 2008; Borgel et al. 2010). Virtually, all these promoters undergo demethylation in developing PGCs, probably from E8.5 (Guibert et al. 2012; Seisenberger et al. 2012). Although most of them control genes that are not expressed in PGCs, a subset are associated with pluripotency and germline-specific factors, suggesting that DNA demethylation could control the germline expression of these specific transcripts (Guibert et al. 2012; Hackett et al. 2012a; Seisenberger et al. 2012; Yamaguchi et al. 2012). The recent identification of a set of germline-specific genes that are apparently primarily regulated by DNA methylation further supports this hypothesis. Consistently, demethylation of their promoters in migratory or post-migratory PGCs leads to gene expression activation. Interestingly, this set of genes, that includes *Tex19.1* and *Mili*, is involved in the genome defence mechanism against parasitic elements (Hackett et al. 2012a). Expression in post-migratory PGCs of other germline-specific genes related to meiosis and germ cell functions is also concomitant with erasure of DNA methylation on their promoters (Maatouk et al. 2006; Seisenberger et al. 2012; Yamaguchi et al. 2012).

Taken together these findings suggest that germline-specific genes are activated by promoter demethylation at different times of PGC development, some during PGC migration (~E8.5) and others upon genital ridge colonisation at ~E11.0 (Maatouk et al. 2006; Guibert et al. 2012; Hackett et al. 2012a; Seisenberger et al. 2012). Similarly, activation of the pluripotency genes is associated with promoter demethylation in migrating PGCs; however their subsequent repression in late GCs (by E16.5) occurs in a DNA methylation-independent manner (Seisenberger et al. 2012). Although less documented, mainly due to technical constraints, histone modifications could also be involved in the control of germline-specific gene expression. For instance, *Dhx38*, which is thought to be repressed by PRMT5-mediated H2A/H4R3me2s in E8.5 PGCs, starts to be expressed at ~E11.5, at the time of the genome-wide removal of this mark (Ancelin et al. 2006).

DNA methylation has a key role in repressing the potentially mutagenic transcriptional activity of transposable elements, which make up about 50 % of the mammalian genome. Most transposable elements are retroelements that can insert into new positions in the genome via a “copy-and-paste” mechanism that involves RNA intermediates (Zamudio and Bourc’his 2010). Like the rest of the genome, these sequences undergo DNA methylation erasure in developing PGCs (Popp et al. 2010; Guibert et al. 2012). The LINE family L1 (long interspersed element 1), for instance, is demethylated in post-migratory PGCs at ~E11.5 (Lane et al. 2003). However, some transposable elements partially resist to demethylation. These include the IAPs and a young subfamily of LTR-ERV1 retroelements (Hajkova et al. 2002; Lane et al. 2003; Popp et al. 2010; Guibert et al. 2012). It is not clear to which extent demethylated retroelements are transcriptionally derepressed. One might expect that to ensure their propagation at an evolutionary scale, transposable elements have to be expressed in GCs. Intriguingly, however, a recent study based on an RNA-seq approach revealed that demethylation is not associated with a general transcriptional activation of L1 family in both female and male GCs by E13.5. A burst of L1 expression is observed later, at E16.5, and

specifically in female GCs (Seisenberger et al. 2012). This observation suggests that mechanisms other than DNA methylation repress L1 expression. Among other possibilities, this repression could act at the post-transcriptional level. Genes involved in the host genome defence mechanism, such as *Tex19.1* and *Mili*, are indeed already active in GCs before E13.5. Transcription of demethylated transposable elements might reveal them to the genome defence mechanism, leading to their repression via post-transcriptional mechanisms before their transcriptional silencing in late GCs (Hackett et al. 2012a) (see Sect. 1.5.3). This system could maintain the genome integrity during germ cell development and would ensure that all active transposable elements are effectively repressed.

1.4 Mechanisms of DNA Demethylation in PGCs

The mechanisms leading to DNA demethylation in mammals are the subject of a long-lasting debate. The simplest mechanism relies on the absence of the DNA methylation machinery. In this so-called passive demethylation process, lack of DNA methylation maintenance during cell replication leads to its progressive dilution over cell divisions. However, DNA demethylation can also occur in a replication-independent manner, indicating that “active” demethylation exists in mammalian cells. Different mechanisms can account for active DNA demethylation, all involving DNA repair through the nucleotide excision repair (NER) or the base excision repair (BER) pathways (reviewed in Wu and Zhang 2010 and Niehrs and Schäfer 2012). NER is a multistep process that relies on several proteins and involves the removal of ~30 bp single-strand DNA sequence that includes the damaged nucleotide. The resulting gap is filled in by DNA polymerases and DNA ligase seals the nicks. The BER machinery is a two-step process in which a specific DNA glycosylase recognises and removes the targeted base from the DNA; gap filling and nick sealing are ensured by DNA polymerases and a DNA ligase.

Passive demethylation probably occurs in migrating PGCs (Fig. 1.3). The transient down-regulation of DNMT1 between E7.5 and E8.25 and the stable repression of its cofactor NP95, which is essential for DNA methylation maintenance (Sharif et al. 2007), probably account for the observed decline in DNA methylation (Seki et al. 2005). However, because most PGCs enter G2 arrest between E7.75 and E8.75, active demethylation might also occur and could involve the growth-arrest and DNA-damage-inducible protein 45 α (GADD45a) that mediates DNA demethylation of specific sequences through the NER or the BER pathway in mammalian and non-mammalian cells (Niehrs and Schäfer 2012; Barreto et al. 2007; Schmitz et al. 2009). *Gadd45a* is up-regulated in PGCs at the time of fate determination and its expression is possibly maintained after G2 arrest (Mochizuki and Matsui 2010; Kurimoto et al. 2008b). However, it is not known whether components of the NER or the BER pathway are also present at this stage. In addition, *Gadd45a*^{-/-} mice do not show defects in fertility (the PGC (de) methylation pattern was not explored in these mutants) (Engel et al. 2009).

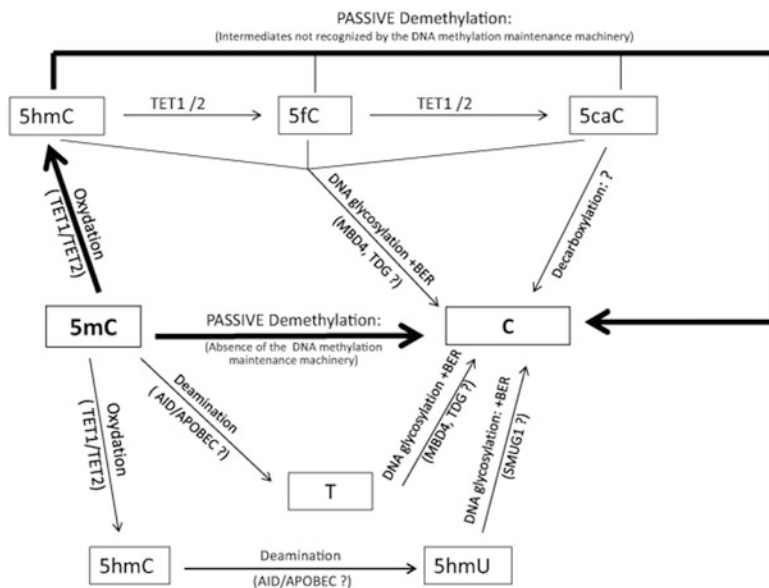


Fig. 1.3 Possible mechanisms of cytosine demethylation during mouse PGC development. During genome-wide PGC DNA demethylation, methylated cytosines (5mC) are converted to cytosines through a passive or an active mechanism, each involving different pathways (see text for details). Absence of the DNA methylation maintenance machinery (i.e., DNMT and cofactors) or its inability to recognise hydroxylated derivatives of 5mC (i.e., 5hmC, 5fC and 5caC) leads to passive dilution of methylation during cell replication. Active replication-independent demethylation in germ cells is believed to involve DNA repair through the base excision repair (BER) pathway. This is initially triggered by deamination of 5mC or 5hmC into thymine or 5hmU, respectively, followed by thymine excision by DNA glycosylase. Hydroxylated derivatives of 5mC can also be the target of active demethylation via excision by DNA glycosylases followed by BER. In the emerging picture (symbolised by *bold arrows*), passive demethylation primarily accounts for the loss of methylation in PGCs. This process is mainly driven by the absence of DNA methylation maintenance in an active cell proliferation context and can further be accentuated at specific loci by a prior Ten-Eleven-Translocation-mediated hydroxylation. C: cytosine, 5mC: 5-methylcytosine, 5caC: 5-carboxylcytosine, 5fC: 5-formylcytosine, 5hmC: 5-hydroxymethylcytosine, 5hmU: 5-hydroxymethyluracil, AID: activation-induced deaminase, APOBEC1: apolipoprotein B mRNA editing enzyme catalytic polypeptide 1, G: guanine, MBD4: methyl CpG-binding domain protein 4, SMUG1: single-strand selective monofunctional uracil DNA glycosylase, T: thymine, TDG: thymine DNA glycosylase, TET: Ten-Eleven-Translocation

Both active (Hajkova et al. 2008, 2010) and passive (Guibert et al. 2012; Seisenberger et al. 2012; Kagiwada et al. 2012; Yamaguchi et al. 2012) processes have been proposed to account for the second DNA demethylation step that takes place after PGC entry in the genital ridges.

The initial observation that this second genome-wide DNA demethylation wave seems to occur rapidly within a single G2 phase suggested that it relies on an active, replication-independent mechanism (Hajkova et al. 2008). In support for such an active DNA demethylation process, several components of the BER, but not the

NER, machinery are active in PGC nuclei at ~E11.5 (Hajkova et al. 2010). BER-mediated DNA demethylation occurs in plants where several glycosylases that recognise and remove 5-methylcytosines have been identified (Gehring and Henikoff 2007). In mammals, however, this class of glycosylases has not been characterised yet. It is thus proposed that 5-methylcytosines are first deaminated into thymines and the resulting T:G mismatches are then targeted by thymine glycosylases, such as MBD4 (methyl CpG-binding domain protein 4) or TDG (thymine-DNA glycosylase), that also activate the BER machinery. AID and APOBEC1, the genes which are expressed in PGCs (Morgan et al. 2004; Hajkova et al. 2010), are possible candidate deaminases. Genetic ablation of *Aid* revealed that this enzyme contributes to DNA demethylation in PGCs (Popp et al. 2010). However, its role is limited, as substantial demethylation still occurs in its absence. At E13.5, female and male *Aid*^{-/-} GCs show 20 and 22 % DNA methylation, respectively, which is higher than the 8 and 16 % observed in wild-type PGCs, but far from the ~75 % observed in somatic neighbours (Popp et al. 2010). *Aid* expression in PGCs is detected only from E12.5 (Hajkova et al. 2010); this also does not support a role for this enzyme in DNA demethylation at E11.5. Moreover, AID- and APOBEC1-mutant mice are viable and fertile (Popp et al. 2010; Morrison et al. 1996). These observations question the involvement of DNA deamination (at least mediated by AID and APOBEC-1) in the massive loss of DNA methylation observed at ~E11.5.

The identification of the ten-eleven-translocation (TET) family of 5mC-dioxygenases provides an alternative mechanism for DNA demethylation. TET proteins (TET1, TET2 and TET3) convert 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) and further catalyse the oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Tahiliani et al. 2009; Ito et al. 2011; He et al. 2011). Several findings support the idea that these modifications are cytosine demethylation intermediates. Specifically, in the zygote, the rapid erasure of DNA methylation from the male pronucleus coincides with a concomitant gain of 5hmC, 5fC and 5caC (Iqbal et al. 2011; Wossidlo et al. 2011; Inoue and Zhang 2011). Maternal depletion of TET3 impairs paternal genome demethylation in preimplantation embryos and leads to neonatal lethality (Guo et al. 2011). A similar pathway could also account for germline DNA demethylation. A recent study indeed revealed that loss of DNA methylation from PGCs at ~E10.5 coincides with a concomitant gain of 5hmC (Hackett et al. 2013), likely mediated by TET1 and TET2 that are expressed in PGCs and peak at this specific stage (Hajkova et al. 2010; Hackett et al. 2013). However, if these two factors are indeed involved in DNA methylation erasure, one might expect that they are functionally redundant (Hackett et al. 2013), as TET1-deficient mice are fertile, although less so than wild type (Dawlaty et al. 2011), and its absence only marginally impairs genome-wide DNA demethylation in E13.5 PGCs (Yamaguchi et al. 2012).

Such TET-mediated demethylation can occur through several pathways (Fig. 1.3). In a deamination-independent mechanism, 5hmC, 5fC and 5caC could be directly recognised and removed by DNA glycosylases, followed by BER. Biochemical approaches have demonstrated that 5fC and 5caC, but not 5hmC, can be removed from DNA by TDG (Maiti and Drohat 2011; Hashimoto et al. 2012).

Deamination could also convert 5hmC into 5 hydroxymethyluracil (5hmU) that can be removed by a DNA glycosylase and repaired by BER. TDG and single-strand-selective monofunctional uracil DNA glycosylase 1 (SMUG1) glycosylate 5hmU in vivo (Guo et al. 2011; Cortellino et al. 2011). However, if one or both these scenarios indeed occur in PGCs, the identity of the involved deaminases and/or glycosylases remains to be determined. As discussed above, AID and APOBEC-1 as well as TDG, which is not detected between E10.5 and E13.5 (Hajkova et al. 2010), are unlikely to account for the massive demethylation at ~E11.5. Further investigations on potential candidates, including other deaminases of the APOBEC family or glycosylases such as MBD4 and SMUG-1, should clarify this issue.

A probably more relevant possibility is that the TET-mediated 5mC derivatives are not recognised by the maintenance methylation machinery, as documented for 5hmC (Valinluck and Sowers 2007), leading to passive demethylation. Recent reports indeed revealed that such a mechanism probably accounts for demethylation of imprinted and meiotic genes (Hackett et al. 2013; Yamaguchi et al. 2012). This is also consistent with other studies supporting that genome-wide DNA demethylation in PGCs occurs primarily via a replication-coupled passive mechanism, mediated by the absence of DNA methylation maintenance (Kagiwada et al. 2012; Seisenberger et al. 2012).

To summarise, the erasure of DNA methylation in developing PGCs is likely to involve multiple passive and active mechanisms, possibly in a locus-specific manner. In the emerging model, a replication-coupled passive mechanism primarily accounts for the loss of methylation in PGCs. This process is mainly driven by the absence of DNA methylation maintenance in an active cell proliferation context and can further be accentuated at specific loci by a prior TET-mediated hydroxylation. In addition, active DNA demethylation, mediated by the BER pathway, could act as an auxiliary mechanism at a limited number of sequences (Fig. 1.3).

1.5 Setting Up New Epigenetic Patterns in Germ Cells

Erasure of the epigenetic pattern in PGCs is completed at ~E12.5. From this “naïve” state, the now so-called GCs start to differentiate into male or female gametes. This process is associated with the acquisition of a new, sex-specific epigenetic profile, as monitored by the dynamic of DNA methylation.

1.5.1 *Targets and Timing of Acquisition of DNA Methylation in Germ Cells*

Mature oocytes and sperms show distinct DNA methylation patterns. Overall, the sperm genome is globally more methylated than the oocyte genome (Howlett and Reik 1991). A genome-wide bisulphite sequencing approach that virtually analysed

all cytosines of the genome showed an average methylation level of ~90 % in sperm cells and of ~40 % in mature oocytes (Kobayashi et al. 2012). This male–female asymmetry is observed also in specific genomic regions. Recent exhaustive maps of cytosine methylation distribution in sperms and oocytes showed that, differently from somatic cells, specific regions, which are called CpG islands (CGIs), are prone to be methylated in germ cells. In somatic mammalian cells, DNA methylation occurs almost exclusively at CpG dinucleotides throughout the genome (Auclair and Weber 2012), but not at CGIs, although they are characterised by high CpG density. These usually short genomic regions (from ~200 bp up to several kb in length) are mostly associated with promoter regions (60 % of the about 23,000 CGIs identified in the mouse genome) (Smallwood and Kelsey 2012) and, as a rule, they are not methylated regardless of the activity of the associated promoter. Conversely, ~1,330 CGIs are specifically methylated in oocytes and ~350 in sperms (Kobayashi et al. 2012), thus representing regions with germline-specific methylation (Borgel et al. 2010; Smallwood et al. 2011; Kobayashi et al. 2012; Smith et al. 2012). This feature was previously described only for ICRs that are characterised as germline-inherited differentially methylated regions (gDMRs) (Arnaud 2010) and are a small minority of the methylated CGIs (about 23 imprinted gDMRs are referenced to date) (Arnaud 2010; Proudhon et al. 2012). Imprinted gDMRs maintain DNA methylation upon fertilisation and throughout development. By this mean, DNA methylation marks the parental origin of the allele in somatic tissues and mediates the mono-allelic expression of the associated imprinted gene(s). By contrast, at most non-imprinted CGIs that are methylated in gametes, DNA methylation is totally or mostly erased during preimplantation reprogramming and they do not display parental-allelic methylation after implantation (Borgel et al. 2010; Smallwood et al. 2011; Kobayashi et al. 2012; Proudhon et al. 2012). The role of gamete-specific CGI methylation remains to be formally established. Transcriptome analysis of oocytes before and after acquisition of DNA methylation failed to show significant differences, questioning the role of CGI methylation in oocyte gene expression (Smallwood et al. 2011). In addition, absence of DNA methylation does not result in infertility per se (Kaneda et al. 2004). As not all CGI-associated DNA methylation is erased at the preimplantation stage, it could affect the level of gene expression in preimplantation embryos and influence first lineage specification (Smallwood et al. 2011; Smallwood and Kelsey 2012).

Besides CGIs, sex-specific DNA methylation is also observed at transposable elements that are highly methylated in sperm and moderately methylated in oocyte. For instance long interspersed elements (LINE) as well as short interspersed elements (SINE), LTR retroelements (long terminal repeat retroelements) and DNA transposons display an average methylation level of 80–85 % in sperm and 40 % in oocyte (Kobayashi et al. 2012).

This DNA methylation landscape is acquired at different times and distinct cellular contexts in paternal and maternal GCs, as exemplified by the dynamics of methylation at imprinted gDMRs and transposable elements (Fig. 1.4). In the male germline, parental imprints and methylation at transposable elements are acquired in

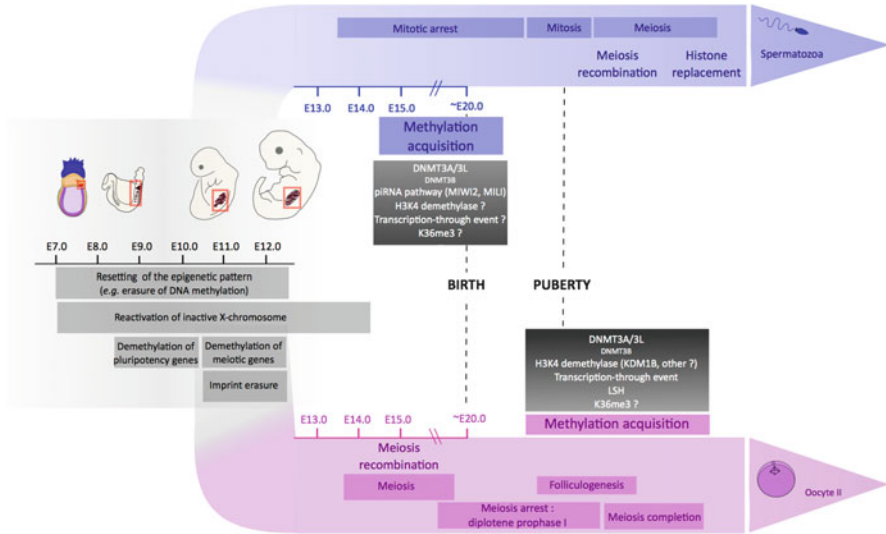


Fig. 1.4 Timing and actors of DNA methylation acquisition in the developing female and male mouse germline. Factors known to be involved in DNA methylation acquisition at single-copy sequences and repetitive elements are shown. The DNA methylation machinery is mainly constituted by DNMT3L and DNMT3A and to a lesser extent by DNMT3B, which is required for methylation of some repetitive element families (see text for details). The main germline developmental stages are indicated

mitotically arrested GCs. Methylation initiates before birth, in pro-spermatogonia at ~E14.5, and is completed in peri-natal pro-spermatogonia and is maintained through the meiotic and haploid stages (Davis et al. 2000; Kato et al. 2007; Ichianagi et al. 2011; Henckel et al. 2011). Conversely, maternal imprints and CGI methylation are acquired after birth in growing oocytes arrested in meiotic prophase I (with a 4n DNA content) (Lucifero et al. 2004; Hiura et al. 2006; Smallwood et al. 2011).

Importantly, the successful completion of meiosis in the male and, to a lesser extent, in the female germline is dependent on the re-establishment of DNA methylation (see Sect. 1.5.3).

1.5.2 Mechanism of Germline DNA Methylation Establishment at Imprinted gDMRs and CGIs

Our knowledge on the mechanism involved in DNA methylation acquisition at imprinted gDMRs has greatly increased in these last years, especially for the gDMRs in the female germline (mat-gDMRs) that account for most of the imprinted gDMRs identified so far (about 20 Mat-gDMRs in the mouse) (Arnaud 2010; Proudhon et al. 2012). Most are promoters and all fulfil the criteria of a CGI,

suggesting that the mechanism involved in their DNA methylation could account also for the methylation of other not imprinted CGIs in the genome (Smallwood and Kelsey 2012). Pat-gDMRs that, unlike mat-gDMR, are CpG-poor non-promoter regions are also treated in this section. Three pat-gDMRs have been characterised (*H19*-DMR, *IG*-DMR and *Rasgrfl*-DMR), while a fourth one, which was initially identified at the *Zdhf2* locus, remains to be firmly validated (Kobayashi et al. 2009, 2012; Proudhon et al. 2012).

The DNA methylation machinery involved in germline de novo methylation is well characterised (Fig. 1.4). A key factor is DNMT3L, a non-catalytic protein that belongs to the de novo methyltransferase 3 family, which includes DNMT3A and DNMT3B. Biochemical studies demonstrated that DNMT3L, by interacting with DNMT3A or B, stimulates their methyltransferase activity (Chedin et al. 2002; Suetake et al. 2004). These three factors are all expressed in the germline, but only DNMT3L and DNMT3A are involved in the methylation of mat-gDMRs and CGIs in oocytes (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004; Smallwood et al. 2011; Kobayashi et al. 2012). In the male germline, DNMT3B also contributes together with DNMT3A to the methylation of pat-gDMRs, as reported for the *Rasgrfl*-DMR (Kato et al. 2007), and these proteins probably catalyse CGI de novo methylation in sperm cells as well.

Many recent studies give some insights into how the DNMT3A/DNMT3L complex targets genomic sequences in a germline-specific manner. The emerging picture is that there is not a universal mechanism as different factors, such as the KRAB zinc finger protein ZFP57, are specifically involved in each germline and/or in a subset of genomic targets (for details see Arnaud 2010; Smallwood and Kelsey 2012). Nonetheless all seem to share a common core pathway with three main features: primary sequence specificity, chromatin configuration and transcriptional events.

The primary sequence and more specifically the spacing between two CpGs in the target sequences have been the focus of much research. Structural analysis showed that the DNMT3A/3L complex preferentially methylates CpGs that are 8–10 bp apart (Jia et al. 2007), suggesting that genomic sequences with this periodicity may be preferential de novo methylation targets. However, in the germline, such a periodicity cannot on its own determine which CGI has to be methylated as it is found at most CGIs, regardless of their methylation status, in mouse oocytes (Ferguson-Smith and Greally 2007; Smallwood et al. 2011). In addition, sequences that do not possess this 8–10 CpG spacing, such as the pat-gDMRs, can also be de novo methylated in the germline.

The chromatin configuration, particularly the histone marks at lysine 4 and 36 of histone H3 (H3K4 and H3K36, respectively), are probably more relevant for germline de novo DNA methylation acquisition. Biochemical approaches have shown that DNMT3L and DNMT3A interact with histone H3 only when it is unmethylated on lysine 4, suggesting that genomic sequences enriched for H3K4me cannot recruit the de novo methylation machinery (Ooi et al. 2007; Zhang et al. 2010). This is in agreement with the observation that DNA methylation and H3K4me2/3 are never associated in mammalian genomes (Barski et al. 2007;

Mikkelsen et al. 2007; Mohn et al. 2008) and that methylated CGIs are depleted of H3K4me3 in oocytes (Smallwood et al. 2011). The hypothesis that H3K4 methylation has to be removed to allow DNA methylation is indirectly supported by functional evidence. In oocytes deficient for KDM1B (an H3K4me1/2 demethylase), a subset of mat-gDMRs failed to acquire DNA methylation (Ciccone et al. 2009). Although DNA methylation at non-imprinted CGIs was not assessed in this study, one can predict that removal of H3K4me, by KDM1B or other H3K4me demethylases, is involved in germline DNA methylation establishment both at imprinted gDMRs and non-imprinted CGIs. Besides H3K4me, H3K36 methylation may also play a role in the de novo methylation mechanism. In vitro, DNMT3A, but not DNMT3B, interacts with H3K36me3 (Dhayalan et al. 2010), suggesting that this mark could recruit the DNMT3A/DNMT3L complex in vivo. Moreover, H3K36me3 association with transcriptionally active regions in mammalian genomes (Barski et al. 2007; Mikkelsen et al. 2007) provides a mechanistic link between transcription and germline DNA methylation establishment. Evidence for this comes from a study conducted on the mouse *Gnas* locus that contains two mat-gDMRs (Coombes et al. 2003). Truncation of *Nesp*, which initiates upstream and overlaps with the two *Gnas* mat-gDMRs, disrupts the acquisition of DNA methylation at these regions in oocytes (Chotalia et al. 2009). Similarly, in the human *GNAS* locus, absence of DNA methylation at the mat-gDMRs correlates with deletion of the *NESP* promoter region (Bastepe et al. 2005). These findings suggest that transcription across gDMRs and CGIs is mechanistically involved in the establishment of their DNA methylation in the germline. Consistent with this model, most mat-gDMRs and CGIs that are methylated in oocytes are located in intragenic regions (Smallwood et al. 2011; Kobayashi et al. 2012; Chotalia et al. 2009). Moreover, transcription across several mat-gDMRs is observed in growing oocytes, when methylation is acquired (Chotalia et al. 2009). Similarly, transcription across the *H19*-DMR and *IG*-DMR is detected in E15.5 and E17.5 pro-spermatogonia, concomitantly with the acquisition of DNA methylation at these pat-gDMRs, suggesting that this mechanism could also play a role in the paternal germline (Henckel et al. 2011). Thus, transcription and the associated deposition of H3K36me3 could “open” the chromatin and facilitate the recruitment of the DNA methylation machinery.

Therefore, in the emerging model, transcriptional read-through event, removal of H3K4me and gain of H3K36me3 act in a concerted manner to recruit DNMT3A and DNMT3L (Fig. 1.5). The exact temporal relationship and interdependence of these events remain to be formally established. For instance, it would be important to determine whether transcriptional read-through and removal of H3K4me are functionally linked or independent events. The observation that in human cells KDM1B is complexed with factors involved in transcription elongation favours the first hypothesis (Fang et al. 2010).

Any model to explain how de novo DNA methylation occurs must also take into account the fact that most CGIs remain unmethylated in the germline. In pro-spermatogonia, when pat-gDMRs acquire their DNA methylation, mat-gDMRs are characterised by active promoters and are enriched for

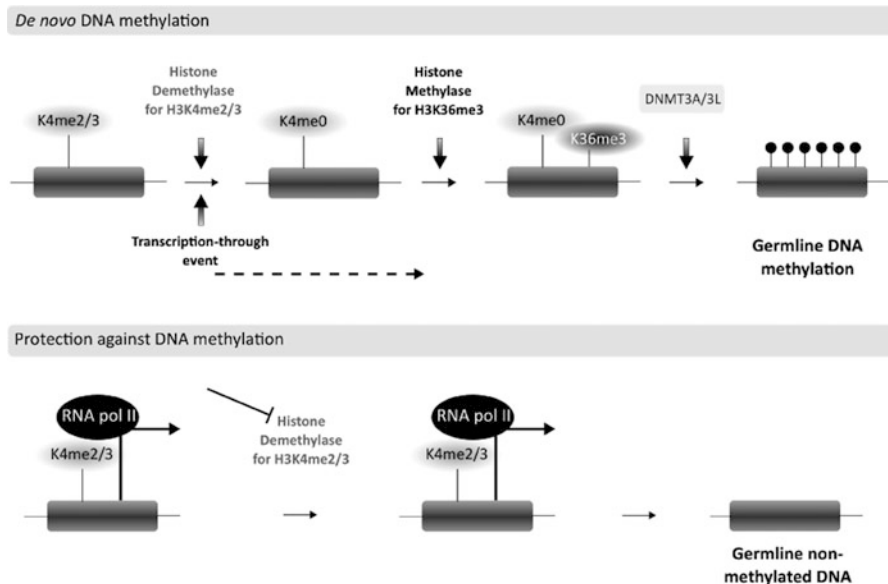


Fig. 1.5 Model of germline DNA methylation establishments. Erasure of H3K4me associated with transcription-through events is required for de novo methylation (*upper part*). In this model H3K36me3 occurs subsequently, but the exact temporal relationship and interdependence between these three events remain to be established. Conversely, promoter activity could prevent H3K4me erasure and recruitment of the DNA methylation machinery (*lower part*). K4me2/3: Di- or trimethylated lysine 4 of histone 3. K4me0: Unmethylated lysine 4 of histone H3. K36me3: Trimethylated lysine 36 of histone 3. RNA polII: RNA polymerase II

H3K4me3 (Henckel et al. 2011). This observation suggests that, in the germline, imprinted gDMRs and CGIs that are also active promoters are protected from DNA methylation (Fig. 1.5). Promoter activity could impair the association of a H3K4 demethylase, thus preventing the recruitment of the DNA methylation machinery.

This model can explain methylation at imprinted and non-imprinted CGIs in oocytes and at the paternally methylated *H19*-DMR and *IG*-DMR. However, methylation in the male germline can also rely on small RNA-based pathway. This, which is well documented at transposable elements (see Sect. 1.5.3), is also observed at the *Rasgrf1*-DMR where acquisition of DNA methylation involves small RNAs and depends on the expression of components of the PIWI-interacting RNA (piRNA) pathway (Watanabe et al. 2011).

1.5.3 Mechanism of Germline DNA Methylation Establishment at Transposable Elements

DNMT3L is the core component of the de novo DNA methylation machinery at transposable elements. In *Dnmt3L*^{-/-} oocytes or pro-spermatogonia, all

transposable element families show reduced DNA methylation (Bourc'his and Bestor 2004; Webster et al. 2005; Kato et al. 2007; Kaneda et al. 2010; Kobayashi et al. 2012). DNMT3A and DNMT3B, which are also involved in this process, have common and differential target specificity. DNMT3A methylates SINE B1 whereas both DNMT3A and B are required for IAP methylation. LINE methylation relies on DNMT3A in the female germline and both DNMT3A and B in the male germline (Kato et al. 2007; Kaneda et al. 2010; Ichihanagi et al. 2011). DNMT3A and/or DNMT3B could also act independently of DNMT3L, particularly at S1NEB1 in the male germline and at retroelements with high CpG density in oocytes (Ichihanagi et al. 2011; Kobayashi et al. 2012).

In the male germline, *Dnmt3L* genetic ablation induced transcriptional reactivation of IAP and LINE retrotransposons and caused severe developmental defects characterised by meiotic failure (possibly due to illegitimate recombination between aberrantly unmethylated non-homologous retrotransposons), progressive loss of germ cells and ultimately arrest of spermatogenesis with complete azoospermia (Bourc'his and Bestor 2004; Webster et al. 2005; Hata et al. 2006). A similar phenotype was also reported in *Dnmt3a* conditional mutant males (Kaneda et al. 2004) and is reminiscent of the defects observed in mutants for components of the piRNA/PIWI pathway (reviewed in Zamudio and Bourc'his 2010).

PiRNAs are a class of small RNAs of ~24–30 nucleotides in length that are specifically expressed in the germline of various vertebrate and invertebrate species (Banisch et al. 2012). They associate with the PIWI subfamily of Argonaute proteins (Girard et al. 2006). MILI and MIWI2, two of the three mouse PIWI family members, are expressed in foetal testes where most piRNAs are derived from retrotransposon elements (Aravin et al. 2006; Kuramochi-Miyagawa et al. 2008). MILI is expressed in male GCs from E12.5 to the round spermatid stage and MIWI2 from E15.5 until birth. In MILI- and MIWI2-deficient foetal testes, a marked decrease in piRNA production correlates with defective de novo methylation and derepression of L1 and IAP elements (Carmell et al. 2007; Kuramochi-Miyagawa et al. 2008; Aravin et al. 2008). Like *Dnmt3L*^{-/-} mice, these mutants cannot complete spermatogenesis and are infertile. A similar phenotype was observed also in mice in which components of the piRNA/PIWI pathway were ablated, such as the tudor domain-containing proteins TDRD1 and TDRD9 or the RNA helicase MVH (mouse vasa homolog) (Zamudio and Bourc'his 2010; Kuramochi-Miyagawa et al. 2010). Altogether, these observations suggest that in foetal testes, retrotransposon-derived piRNAs play a pivotal role in de novo DNA methylation and silencing. According to the current working model, following erasure of DNA methylation, retroelements are expressed from dispersed loci in late PGCs. The occurrence of bidirectional transcription across some loci would lead to the production of both sense and antisense retrotransposon-containing transcripts. These transcripts will be sensed by the PIWI proteins and cleaved into primary sense and secondary antisense piRNAs, which will associate, respectively, with MILI and MIWI2 in distinct cytoplasmic compartments. Exchange of piRNAs between these cytoplasmic compartments triggers an auto-amplification

process that increases the pool of sense and antisense piRNAs and might lead to post-transcriptional silencing of retrotransposons. Translocation of MIWI2 and the associated antisense piRNAs to the nucleus promote de novo DNA methylation of complementary genomic copies of retrotransposons. The underlying molecular mechanism remains unclear. Interactome approaches failed to reveal a physical interaction between MIWI2 and the DNA methylation machinery, suggesting that an intermediate step is required to recruit DNMTs (Castañeda et al. 2011). As in the model of de novo methylation at single copy CGI sequences (Fig. 1.5), the MIWI2/secondary piRNA complex might alter the histone modification signature at retrotransposons, for instance by recruiting an H3K4me demethylase that subsequently facilitates the recruitment of the DNA methylation machinery.

The mechanism that controls transposable element DNA methylation in the female germline is less well characterised. It is not known whether alteration of DNA methylation at transposable elements in *Dnmt3a*^{-/-} or *Dnmt3l*^{-/-} oocytes is associated with increased transcription. Nonetheless these oocytes are apparently normal and support fertilisation (although the derived embryos die in utero) (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004; Kobayashi et al. 2012). PiRNAs and small interfering RNAs (siRNAs), these latter being specifically derived from transposable elements, are also abundant in developing mouse oocytes at the time of DNA methylation establishment (Watanabe et al. 2006, 2008, 2011). However, transposable element expression level is barely affected in oocytes deficient for MILI and DICER and *Mili*^{-/-} and *Miwi2*^{-/-} female mice are fertile (Watanabe et al. 2006; Shoji et al. 2009). Although it is not known whether in these mutants DNA methylation at transposable elements is affected, these observations suggest that in the female germline the RNA-based pathway is not involved in the acquisition of a repressive signature at transposable elements. Indeed, a study that focused on lymphoid-specific helicase (LSH) rather supports a chromatin-based mechanism. This member of the SNF2-helicase family of chromatin remodelling proteins is involved in DNA methylation and transcriptional repression of transposable elements in mammalian somatic cells (Dennis et al. 2001). Interestingly, *Lsh*^{-/-} female mice also show demethylation of IAP elements in pachytene oocytes, increased expression of such elements in *Lsh*^{-/-} ovaries and methylation defects in pericentromeric satellite repeats (De la Fuente et al. 2006). This phenotype is associated with incomplete synapsis of homologous chromosomes and severe oocyte loss in the early postnatal stages, suggesting that LSH is required to complete female meiosis (De la Fuente et al. 2006). Whether LSH also controls methylation at other transposable elements and whether it is involved in the acquisition and/or maintenance of DNA methylation in the female germline is not known.

The sex-specific differences in DNA methylation establishment at transposable elements could be explained by the fact that the female germline might tolerate the presence of transposable element-derived transcripts and thus their repression is not so crucial (Aravin and Bourc'his 2008; Zamudio and Bourc'his 2010). Indeed, as active cell division facilitates (retro-)transposition events, they are thus less likely to occur in cell cycle-arrested growing oocytes.

1.5.4 Chromatin Changes During Gamete Maturation

During the last step of maturation, gametes undergo developmental changes to sustain fertilisation and the first stages of zygotic development (Sasaki and Matsui 2008). This is predominantly observed in the male germline where post-meiotic spermiogenesis is associated with global chromatin remodelling (reviewed in Kota and Feil 2010). The main outcome of this process is the exchange of canonical histones for protamines (small basic proteins), resulting in tightly condensed sperm DNA. The compacted genome in the sperm head is believed to facilitate its motility and protect from DNA damage (Jenkins and Carrell 2012). This event questioned whether and how male gamete could transmit their epigenetic information encoded in histone and associated modifications. The answer came with the finding that in human mature spermatozoa, 5–15 % of their genome remains nucleosome bound (about 1 % in the mouse) (Wykes and Krawetz 2003; Hammoud et al. 2009; Bryczynska et al. 2010). Histone retention is not random and it is found, for instance, at the regulatory regions of key developmental genes, including imprinted genes, microRNAs and homeotic genes. Interestingly, the promoters of genes involved in spermatogenesis and cell homeostasis are enriched for H3K4me2/3 (permissive histone modifications), in agreement with their activation during gametogenesis. In contrast, key genes for embryonic development or morphogenesis harbour the repressive H3K27me3 and eventually, the permissive H3K4me2 marks. These so-called bivalent chromatin domains, initially identified in embryonic stem cells, are believed to poise genes for either activation or repression later in development (Bernstein et al. 2006). Altogether, this suggests that male gametes can convey instructive epigenetic information to the zygote, which can subsequently regulate expression of key embryonic developmental genes.

Alterations in histone retention associated with moderate changes in the amount of H3K4me2/3 and H3K27me3 at some developmental genes and imprinted loci have been detected in sperm of infertile men. These defects were also associated with DNA methylation alterations, though to a limited extent (Hammoud et al. 2011), as previously reported for imprinted genes (Marques et al. 2004, 2008; Kobayashi et al. 2007). These findings suggest that subtle, naturally occurring changes in the chromatin signature of spermatozoa could account for the intra- and inter-individual differences in DNA methylation detected in human sperm samples (Flanagan et al. 2006) and might thus potentially contribute to the progeny phenotypic differences.

1.6 Escaping the Germline Epigenetic Reprogramming

Germline reprogramming affects all the epigenetic-based developmental processes. During PGC development, the epigenetic program of epiblast cells is fully erased and at ~E12.5, the germ cell genome reaches a basal (virtually epigenetic-free) state.

In addition to prevent that epimutations pass on to the next generation, this event is also required for the totipotency potential of the future gametes (Hackett et al. 2012b). The epigenetic-naïve germ cell genome will then acquire a germline-specific epigenetic signature that, in addition to be important for gametogenesis itself, can be transmitted to and interpreted by the progeny. This is illustrated, for instance, by genomic imprinting. The paternal and maternal imprints inherited from epiblast cells are erased in PGCs and a new set of sex-specific imprints is acquired by the developing germlines. This epigenetic signature is faithfully maintained following fertilisation and is transmitted and interpreted by the somatic cell lineages of the new individual.

Besides this well-characterised form of transmission of epigenetic information from one generation to the next one, the recurrent observation that some transposable elements can escape DNA methylation reprogramming suggests that epigenetic inheritance might also occur across several generations. IAPs and the subfamily of LTR-ERV1 retroelements, which have been shown to resist both germline and preimplantation reprogramming, are the prototype of such transgenerational inheritance of an epigenetic state that could affect also neighbouring sequences (Hajkova et al. 2002; Lane et al. 2003; Popp et al. 2010; Guibert et al. 2012; Kobayashi et al. 2012). This is in line with the few known cases of transgenerational epimutation in mammals that involve IAP elements and can alter the expression of neighbouring genes in a variegated manner (Daxinger and Whitelaw 2012).

Besides transposable element, recent genome-wide analyses identified few single-copy sequences that resist both waves of DNA methylation erasure in PGCs and preimplantation embryos (Smallwood et al. 2011; Guibert et al. 2012). In a pioneer study about 20 of such sequences, including intergenic, intragenic and promoter regions, have been formally identified (Guibert et al. 2012) and others remain to be revealed (Seisenberger et al. 2012; Hackett et al. 2013); however, their function (if any), particularly during embryo development, remains to be established. Nonetheless these observations indicate that the DNA methylation pattern at single-copy sequences can be transmitted across generations.

Several lines of evidence suggest that besides DNA methylation, the epigenetic signature is not fully erased in PGCs. Indeed, in the developing germline, the timing of methylation acquisition at imprinted gDMRs can be different at the two alleles according to their parental origin. For instance, in male germ cells, methylation at the *H19*-DMR is first reacquired on the paternally inherited allele (Davis et al. 2000). Similarly, in the female germline several mat-gDMRs first acquire methylation on the maternally inherited allele (Lucifero et al. 2004; Hiura et al. 2006). This suggests that following DNA methylation erasure, alleles can nevertheless remember their parental origin, possibly through other epigenetic modifications. As a mirror event, it is also documented that defects in the acquisition of germline methylation at imprinted mat-gDMRs can be partially rescued during embryonic development. This is mainly observed at the *Snrpn*-DMR in embryos derived from *Dnmt3L*^{-/-} or *Zfp57*^{-/-} oocytes. Although DNMT3L and ZFP57 are crucial for its oocyte methylation, in some of these embryos

the *Snrpn*-DMR is methylated on the maternal allele (Arnaud et al. 2006; Li et al. 2008). This suggests that a yet-to-be-identified germline-derived signature escapes preimplantation reprogramming and mediates parental allele-specific DNA methylation acquisition during early embryonic development.

Although these observations are not associated with transgenerational epigenetic inheritance per se, they provide the proof of concept that an epigenetic DNA methylation-independent signature could be transmitted across generations.

1.7 Conclusions

These last years witnessed major breakthroughs in our understanding of the mechanisms involved in germline reprogramming, especially on how the new epigenetic signature is acquired. The emergence of dedicated deep-sequencing technologies that allowed the generation of an exhaustive map of DNA methylation from a limited amount of genomic DNA has been instrumental in these advances. The use of these technologies to assess migrating and early-post-migrating PGCs, combined with the recently developed genome-wide approaches to profile 5hmC (Schüler and Miller 2012), should provide, in the very near future, key information on the kinetics, extent and mechanisms of germline DNA methylation erasure. These data will be important for distinguishing between passive and active demethylation mechanisms, establishing the role of hydroxylated 5mC derivatives in this process and revealing the nature and identity of all the genomic sequences that resist erasure. However, DNA methylation is not the only component of the epigenetic signature. Study of germline reprogramming should also rely on the optimisation of “scaled-down” approaches to determine in vivo the dynamics of histone modifications (by Chip-seq) and transcription events (by RNA-seq) in the developing germline. Besides this highly demanding and challenging in vivo approaches, the development of cell culture models should provide a versatile tool to analyse the molecular events of germ cell development and reprogramming. The improvement of the already existing in vitro or ex vivo models towards a model that faithfully recapitulates germ cell development will provide a powerful controlled system to monitor all reprogramming steps (Gillich and Hayashi 2011).

These efforts should lead to the identification of the key factors involved in the resetting and (re)acquisition of the epigenetic pattern. Besides their fundamental importance for our understanding of epigenetic inheritance, this knowledge will have major clinical impacts with potential applications in reproductive and regenerative medicine. In addition, it will help tackling the fascinating question of whether the environment can influence reprogramming between generations. As most of the germline reprogramming occurs during embryonic life in the womb, one might speculate that environmental factors, to which mothers are exposed during pregnancy, could affect the phenotype of not only her children but also her grandchildren.

Acknowledgements We would like to thank Dr Claire Chazaud for help with the figure design and Dr Yoan Renaud for daily discussions. The authors' team is financially supported by a French National Research Agency (ANR: project n° 2010 JCJC 1601), a "La Ligue- comité du Puy-de-Dôme" and a "Conseil régional d'Auvergne" grants awarded to PA. SMM is a recipient of a "Fondation ARC pour la Recherche sur le Cancer" fellowship.

References

- Ancelin K, Lange UC, Hajkova P, Schneider R, Bannister AJ, Kouzarides T, Surani MA (2006) *Blimp1* associates with *Prmt5* and directs histone arginine methylation in mouse germ cells. *Nat Cell Biol* 8:623–630
- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442:203–207
- Aravin AA, Bourc'his D (2008) Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes Dev* 22:970–975
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31:785–799
- Arnaud P, Hata K, Kaneda M, Li E, Sasaki H, Feil R, Kelsey G (2006) Stochastic imprinting in the progeny of *Dnmt3L*^{-/-} females. *Hum Mol Genet* 15:589–598
- Arnaud P (2010) Genomic imprinting in germ cells: imprints are under control. *Reproduction* 140:411–423
- Auclair G, Weber M (2012) Mechanisms of DNA methylation and demethylation in mammals. *Biochimie* 94(11):2202–2211. doi:[10.1016/j.biochi.2012.05.016](https://doi.org/10.1016/j.biochi.2012.05.016)
- Banisch TU, Goudarzi M, Raz E (2012) Small RNAs in germ cell development. *Curr Top Dev Biol* 99:79–113
- Barreto G, Schäfer A, Marhold J, Stach D, Swaminathan SK, Handa V, Döderlein G, Maltry N, Wu W, Lyko F, Niehrs C (2007) *Gadd45a* promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* 445:671–675
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837
- Bastepe M, Fröhlich LF, Linglart A, Abu-Zahra HS, Tojo K, Ward LM, Jüppner H (2005) Deletion of the NESP55 differentially methylated region causes loss of maternal *GNAS* imprints and pseudohypoparathyroidism type 1b. *Nat Genet* 37:25–27
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125:315–326
- Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42:1093–1100
- Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking *Dnmt3L*. *Nature* 431:96–99
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) *Dnmt3L* and the establishment of maternal genomic imprints. *Science* 294:2536–2539
- Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, Beisel C, Schübeler D, Stadler MB, Peters AH (2010) Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* 17:679–687

- Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 12:503–514
- Castañeda J, Genzor P, Bortvin A (2011) piRNAs, transposon silencing, and germline genome integrity. *Mutat Res* 714:95–104
- Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, Robertson M, Vrana J, Jones K, Grotewold L, Smith A (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* 450:1230–1234
- Chedin F, Lieber MR, Hsieh CL (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* 99:16916–16921
- Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev* 23:105–117
- Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461:415–418
- Coombes C, Arnaud P, Gordon E, Dean W, Coar EA, Williamson CM, Feil R, Peters J, Kelsey G (2003) Epigenetic properties and identification of an imprint mark in the Nesp-Gnasxl domain of the mouse Gnas imprinted locus. *Mol Cell Biol* 23:5475–5488
- Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146:67–79
- Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C, Zeissler U, Lappe M, Schübeler D, Torres-Padilla ME, Schneider R (2009) H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nat Struct Mol Biol* 16:777–781
- Davis TL, Yang GJ, McCarrey JR, Bartolomei MS (2000) The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet* 9:2885–2894
- Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, Gao Q, Kim J, Choi SW, Page DC, Jaenisch R (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9:166–175
- Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* 13:153–162. doi:[10.1038/nrg3188](https://doi.org/10.1038/nrg3188)
- De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, Muegge K (2006) Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nat Cell Biol* 8:1448–1454
- De Napolés M, Nesterova T, Brockdorff N (2007) Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells. *PLoS One* 2(9):e860
- Dennis K, Fan T, Geiman T, Yan Q, Muegge K (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* 15:2940–2944
- Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* 285:26114–26120
- Engel N, Tront JS, Erinle T, Nguyen N, Latham KE, Sapienza C, Hoffman B, Liebermann DA (2009) Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics* 4:98–99
- Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C, Peng J, Cheng D, Sui G, Kaiser UB, Shi Y, Shi YG (2010) Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. *Mol Cell* 39:222–233
- Ferguson-Smith AC, Greally JM (2007) Epigenetics: perceptive enzymes. *Nature* 449:148–149

- Flanagan JM, Popenkityte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, Schumacher A, Zangeneh M, Lau L, Virtanen C, Wang SC, Petronis A (2006) Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* 79:67–84
- Gehring M, Henikoff S (2007) DNA methylation dynamics in plant genomes. *Biochim Biophys Acta* 1769:276–286
- Gillich A, Hayashi K (2011) Switching stem cell state through programmed germ cell reprogramming. *Differentiation* 8:281–291
- Ginsburg M, Snow MH, McLaren A (1990) Primordial germ cells in the mouse embryo during gastrulation. *Development* 110:521–528
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202
- Guibert S, Forné T, Weber M (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* 22:633–641
- Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle* 10:2662–2668
- Gyory I, Wu J, Fejér G, Seto E, Wright KL (2004) PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing. *Nat Immunol* 5:299–308
- Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, Reik W, Surani MA, Adams IR, Meehan RR (2012a) Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* 139:3623–3632. doi:[10.1242/dev.081661](https://doi.org/10.1242/dev.081661)
- Hackett JA, Zyllicz JJ, Surani MA (2012b) Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet* 28:164–174
- Hackett JA, Sengupta R, Zyllicz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339:448–452. doi:[10.1126/science.1229277](https://doi.org/10.1126/science.1229277)
- Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA (2008) Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 452:877–881
- Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 117:15–23
- Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA (2010) Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329:78–82
- Hammoud SS, Nix DA, Hammoud AO, Gibson M, Cairns BR, Carrell DT (2011) Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Hum Reprod* 26:2558–2569
- Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460:473–478
- Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X (2012) Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. *Nucleic Acids Res* 40:10203–10214. doi:[10.1093/nar/gks845](https://doi.org/10.1093/nar/gks845)
- Hata K, Kusumi M, Yokomine T, Li E, Sasaki H (2006) Meiotic and epigenetic aberrations in Dnmt3L-deficient male germ cells. *Mol Reprod Dev* 73:116–122
- Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129:1983–1993
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–1307
- Henckel A, Chebli K, Kota SK, Arnaud P, Feil R (2011) Transcription and histone methylation changes correlate with imprint acquisition in male germ cells. *EMBO J* 31:606–615. doi:[10.1038/emboj.2011.425](https://doi.org/10.1038/emboj.2011.425)

- Hiura H, Obata Y, Komiyama J, Shirai M, Kono T (2006) Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* 11:353–361
- Howlett SK, Reik W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113:119–127
- Ichiiyanagi K, Li Y, Watanabe T, Ichiiyanagi T, Fukuda K, Kitayama J, Yamamoto Y, Kuramochi-Miyagawa S, Nakano T, Yabuta Y, Seki Y, Saitou M, Sasaki H (2011) Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. *Genome Res* 21:2058–2066
- Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334:194
- Iqbal K, Jin SG, Pfeifer GP, Szabó PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 108:3642–3647
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–1333
- Jenkins TG, Carrell DT (2012) The sperm epigenome and potential implications for the developing embryo. *Reproduction* 143:727–734
- Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449:248–251
- Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M (2012) Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J* (Advance online publication, 14 December 2012; doi:10.1038/emboj.2012.331)
- Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, Sasaki H (2010) Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. *Genes Cells* 15:169–179. doi:10.1111/j.1365-2443.2009.01374.x
- Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429:900–903
- Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* 16:2272–2280
- Kato Y, Rideout WM 3rd, Hilton K, Barton SC, Tsunoda Y, Surani MA (1999) Developmental potential of mouse primordial germ cells. *Development* 126:1823–1832
- Kehler J, Tolkunova E, Koschorz B, Pesce M, Gentile L, Boiani M, Lomelí H, Nagy A, McLaughlin KJ, Schöler HR, Tomilin A (2004) Oct4 is required for primordial germ cell survival. *EMBO Rep* 5:1078–1083
- Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T (2012) Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet* 8 (1):e1002440. doi:10.1371/journal.pgen.1002440
- Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, Sasaki H, Yaegashi N, Arima T (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16:2542–2551
- Kobayashi H, Yamada K, Morita S, Hiura H, Fukuda A, Kagami M, Ogata T, Hata K, Sotomaru Y, Kono T (2009) Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. *Genomics* 93:461–472
- Kota SK, Feil R (2010) Epigenetic transitions in germ cell development and meiosis. *Dev Cell* 19:675–686
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Takamatsu K, Chuma S, Kojima-Kita K, Shiromoto Y, Asada N, Toyoda A, Fujiyama A, Totoki Y, Shibata T, Kimura T,

- Nakatsuji N, Noce T, Sasaki H, Nakano T (2010) MVH in piRNA processing and gene silencing of retrotransposons. *Genes Dev* 24:887–892
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 22:908–917
- Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M (2008a) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev* 22:1617–1635
- Kurimoto K, Yamaji M, Seki Y, Saitou M (2008b) Specification of the germ cell lineage in mice: a process orchestrated by the PR-domain proteins, Blimp1 and Prdm14. *Cell Cycle* 7(22):3514–3518
- Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35:88–93
- Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, Korving JP, Hogan BL (1999) Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* 13:424–436
- Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, Ogura A, Ishino F (2002) Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* 129:1807–1817
- Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC (2008) A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell* 15:547–557
- Lucifero D, Mann MR, Bartolomei MS, Trasler JM (2004) Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum Mol Genet* 13:839–849
- Maatouk DM, Kellam LD, Mann MR, Lei H, Li E, Bartolomei MS, Resnick JL (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* 133:3411–3418
- Maiti A, Drohat AC (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286:35334–35338. doi:[10.1074/jbc.C111.284620](https://doi.org/10.1074/jbc.C111.284620)
- Mansour AA, Gafni O, Weinberger L, Zviran A, Ayyash M, Rais Y, Krupalnik V, Zerbib M, Amann-Zalcenstein D, Maza I, Geula S, Viukov S, Holtzman L, Pribluda A, Canaani E, Horn-Saban S, Amit I, Novershtern N, Hanna JH (2012) The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* 488:409–4013
- Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. *Lancet* 363:1700–1702
- Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, Sousa M (2008) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14:67–74
- Matsui Y, Zsebo K, Hogan BL (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 70:841–847
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560
- Mochizuki K, Matsui Y (2010) Epigenetic profiles in primordial germ cells: global modulation and fine tuning of the epigenome for acquisition of totipotency. *Dev Growth Differ* 52:517–525
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30:755–766

- Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK (2004) Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* 279:52353–52360
- Morrison JR, Pászty C, Stevens ME, Hughes SD, Forte T, Scott J, Rubin EM (1996) Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc Natl Acad Sci U S A* 93:7154–7159
- Niehrs C, Schäfer A (2012) Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol* 22:220–227
- Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M (2009) A signaling principle for the specification of the germ cell lineage in mice. *Cell* 137:571–584
- Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, Barton SC, Obukhanych T, Nussenzweig M, Tarakhovsky A, Saitou M, Surani MA (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* 436:207–213
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448:714–717
- Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463:1101–1105
- Proudhon C, Duffié R, Ajjan S, Cowley M, Iranzo J, Carbajosa G, Saadeh H, Holland ML, Oakey RJ, Rakyanc VK, Schulz R, Bourc'his D (2012) Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol Cell* 47:909–920. doi:[10.1016/j.molcel.2012.07.010](https://doi.org/10.1016/j.molcel.2012.07.010)
- Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development reprogramming and beyond. *Nat Rev Genet* 9:129–140
- Schmitz KM, Schmitt N, Hoffmann-Rohrer U, Schäfer A, Grummt I, Mayer C (2009) TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol Cell* 233:344–353
- Schüler P, Miller AK (2012) Sequencing the sixth base (5-hydroxymethylcytosine): selective DNA oxidation enables base-pair resolution. *Angew Chem Int Ed Engl* 51:10704–10707. doi:[10.1002/anie.201204768](https://doi.org/10.1002/anie.201204768)
- Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W (2012) The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* 48(6):849–862. doi:[10.1016/j.molcel.2012.11.001](https://doi.org/10.1016/j.molcel.2012.11.001)
- Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y (2005) Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev Biol* 278:440–458
- Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* 134:2627–2638
- Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450:908–912
- Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y, Kondoh G, Okawa K, Chujo T, Suzuki T, Hata K, Martin SL, Noce T, Kuramochi-Miyagawa S, Nakano T, Sasaki H, Pillai RS, Nakatsuji N, Chuma S (2009) The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev Cell* 17:775–787
- Smallwood SA, Kelsey G (2012) De novo DNA methylation: a germ cell perspective. *Trends Genet* 28:33–42
- Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* 43:811–814. doi:[10.1038/ng.864](https://doi.org/10.1038/ng.864)

- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484:339–344. doi:10.1038/nature10960
- Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S (2004) DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 279:27816–27823
- Sugimoto M, Abe K (2007) X chromosome reactivation initiates in nascent primordial germ cells in mice. *PLoS Genet* 3(7):e116
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935
- Tam PP, Zhou SX, Tan SS (1994) X-chromosome activity of the mouse primordial germ cells revealed by the expression of an X-linked lacZ transgene. *Development* 120:2925–2932
- Valinluck V, Sowers LC (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 67:946–950
- Vincent SD, Dunn NR, Sciammas R, Shapiro-Shalef M, Davis MM, Calame K, Bikoff EK, Robertson EJ (2005) The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* 132:1315–1325
- Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H (2006) Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* 20:1732–1743
- Watanabe T, Tomizawa S, Mitsuya K, Totoki Y, Yamamoto Y, Kuramochi-Miyagawa S, Iida N, Hoki Y, Murphy PJ, Toyoda A, Gotoh K, Hiura H, Arima T, Fujiyama A, Sado T, Shibata T, Nakano T, Lin H, Ichiyangi K, Soloway PD, Sasaki H (2011) Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse *Rasgrf1* locus. *Science* 332:848–852
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453:539–543
- Weber M, Schübeler D (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* 19:273–280
- Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, Harrison DK, Aung H, Phutikanit N, Lyle R, Meachem SJ, Antonarakis SE, de Kretser DM, Hedger MP, Peterson P, Carroll BJ, Scott HS (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc Natl Acad Sci U S A* 102:4068–4073
- West JA, Viswanathan SR, Yabuuchi A, Cunniff K, Takeuchi A, Park IH, Sero JE, Zhu H, Perez-Atayde A, Frazier AL, Surani MA, Daley GQ (2009) A role for Lin28 in primordial germ-cell development and germ-cell malignancy. *Nature* 460:909–913
- Western PS, Miles DC, van den Bergen JA, Burton M, Sinclair AH (2008) Dynamic regulation of mitotic arrest in fetal male germ cells. *Stem Cells* 26:339–347
- Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2:241
- Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 11:607–620
- Wykes SM, Krawetz SA (2003) The structural organization of sperm chromatin. *J Biol Chem* 278:29471–29477
- Yamagata K, Yamazaki T, Miki H, Ogonuki N, Inoue K, Ogura A, Baba T (2007) Centromeric DNA hypomethylation as an epigenetic signature discriminates between germ and somatic cell lineages. *Dev Biol* 312:419–426

- Yamaguchi S, Hong K, Liu R, Shen L, Inoue A, Diep D, Zhang K, Zhang Y (2012) Tet1 controls meiosis by regulating meiotic gene expression. *Nature* 492(7429):443–447. doi:[10.1038/nature11709](https://doi.org/10.1038/nature11709)
- Yamaguchi S, Kurimoto K, Yabuta Y, Sasaki H, Nakatsuji N, Saitou M, Tada T (2009) Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. *Development* 136:4011–4020
- Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, Yamanaka K, Ohinata Y, Saitou M (2008) Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat Genet* 40:1016–1022
- Yeom YI, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, Hübner K, Schöler HR (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* 122:881–894
- Yu J, Angelin-Duclos C, Greenwood J, Liao J, Calame K (2000) Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol Cell Biol* 20:2592–2603
- Zamudio N, Bourc'his D (2010) Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity (Edinb)* 105:92–104
- Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* 38:4246–4253

Chapter 2

Establishment of Tissue-Specific Epigenetic States During Development

Ionel Sandovici

Abstract Complex organisms require tissue-specific transcriptional programs, which are acquired during development through the stepwise activation of transcription factor networks acting in tight coordination with epigenetic mechanisms. Recent progresses in genome-wide mapping of various epigenetic marks across a panel of mammalian cell types and developmental stages, together with a multitude of functional analyses, led to significant advances in our understanding of tissue-specific epigenetic regulation of gene expression. These new developments open at last the opportunity to systematically explore the contribution of epigenetics to human disease.

2.1 Introduction

Complex organisms such as mammals comprise over 200 different cell types, each one expressing specific sets of genes that define their unique functions (Alberts et al. 2002). The tissue-specific transcriptional programs are acquired over the course of development, during which cells transit from a pluripotent state to differentiated cell lineages, in a well-orchestrated spatial and temporal manner. This stepwise process is controlled by the sequential activation of specific transcription factors acting coordinately with epigenetic mechanisms (Reik et al. 2001; Hemberger et al. 2009; Albert and Peters 2009), which particularly target key DNA regulatory sequences such as promoters, enhancers, and insulators (Maston et al. 2006).

I. Sandovici (✉)

Metabolic Research Laboratories, Department of Obstetrics and Gynaecology,
University of Cambridge, Cambridge CB2 0SW, UK

Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 3EG, UK
e-mail: ionel.sandovici@bbsrc.ac.uk

Epigenetics refers to heritable changes in gene expression, caused by mechanisms other than changes in the underlying DNA sequence (Bird 2007). These heritable changes in gene expression are brought about by a complex array of reversible epigenetic marks: DNA modifications (such as 5-methylcytosine (Suzuki and Bird 2008), 5-hydroxymethylcytosine (Tahiliani et al. 2009), 5-formylcytosine, and 5-carboxylcytosine (Ito et al. 2011)), posttranslational modifications of histones (such as methylation, acetylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, deimination, proline isomerization) (Bannister and Kouzarides 2011), histone variants (Talbert and Henikoff 2010), and alternative nucleosome positioning (Bai and Morozov 2010). The epigenetic marks are laid on the chromatin by an array of chromatin- and DNA-binding proteins with enzymatic activities (Brenner and Fuks 2007), as well as noncoding RNAs (Grewal 2010), all of which act as epigenetic initiators (epigenators) (Berger et al. 2009).

In the past few years significant advances in our understanding of tissue-specific epigenetic regulation of gene expression have been made possible by loss-of-function studies, as well as genome-wide mapping of different epigenetic marks across a panel of mammalian cell types and developmental stages. A seminal contribution in this direction has been provided by the Encyclopedia of DNA Elements (ENCODE) project. Very recently, the human ENCODE project has achieved the systematic characterization of a large variety of epigenetic marks in 147 different cell types (ENCODE Project Consortium 2012). Moreover, the integration of ENCODE data with other resources such as the genome-wide association studies (GWAS) has started to reveal new insights into human disease (Maurano et al. 2012). In this chapter I summarize the current view on the establishment of tissue-specific epigenetic marks during development, how these epigenetic patterns are correlated with transcription in a cell-specific manner, and how the tissue-specific epigenetic states may be directly linked with human disease.

2.2 Epigenetic Reprogramming During Preimplantation Development

The life of an organism begins at fertilization with the formation of the zygote. Fertilization coincides with a wave of epigenetic reprogramming (Fig. 2.1) that is required for the achievement of developmental totipotency (Reik et al. 2001; Hemberger et al. 2009; Albert and Peters 2009). Interestingly, even before fertilization the oocyte exhibits global hypomethylation, particularly at specific families of long interspersed element 1 (LINE1) and long terminal repeat (LTR) retroelements (Smith et al. 2012). A major initial event in the post-fertilization reprogramming process is the active loss of DNA methylation in the paternal pronucleus (Santos et al. 2002), likely by partial conversion of 5-methylcytosine into 5-hydroxymethylcytosine by TET3 (ten–eleven translocation) protein (Wossidlo et al. 2011; Gu et al. 2011). This is followed by passive DNA

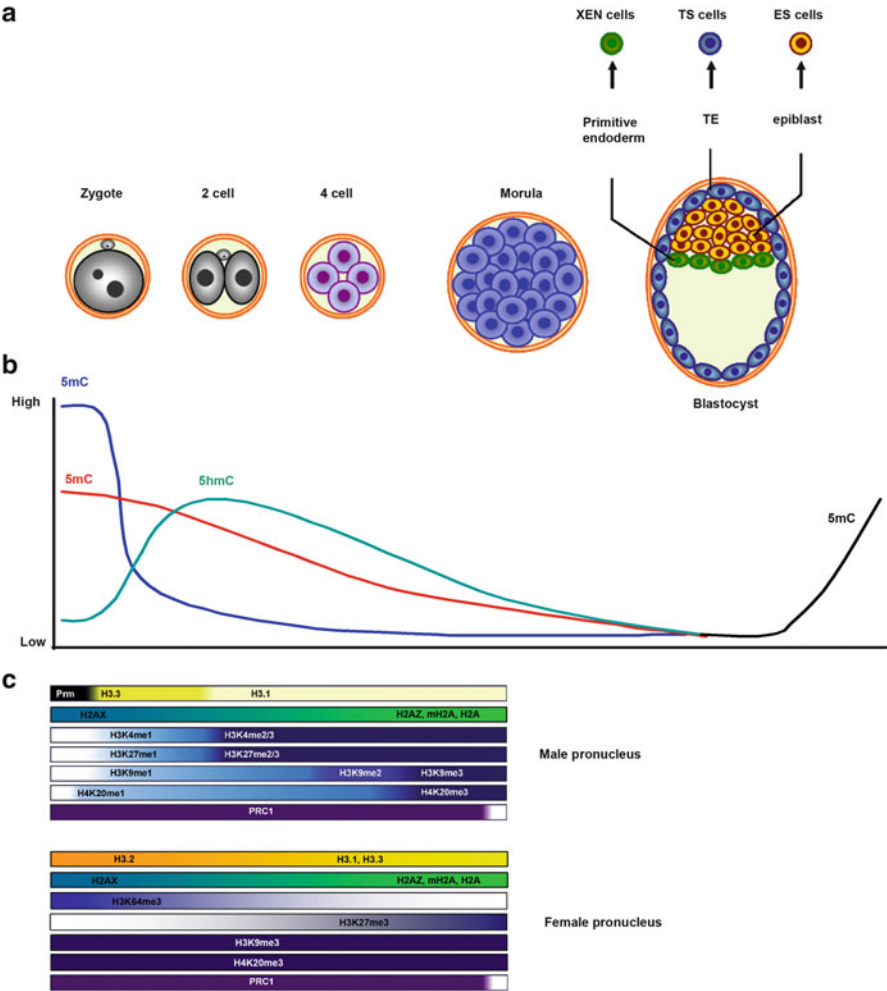


Fig. 2.1 Epigenetic reprogramming in early embryos (see text for details). (a) Diagram of the first events during the preimplantation development. XEN—extraembryonic endoderm stem cells, ES—embryonic stem cells, TE—trophoblast, TS—trophoblast stem cells. (b) Dynamics of DNA modification changes during early development: blue—paternal pronucleus; red—maternal pronucleus. 5mC—5-methylcytosine, 5hmC—5-hydroxymethylcytosine. (c) Histone variants and histone modification dynamics during early development (adapted from Hemberger et al. (2009); Albert and Peters (2009)). Prm—protamines; H3.1, H3.2, H3.3—histone H3 variants; H2AX, H2AZ, mH2A, and H2A—histone H2 variants; K—lysine; me—methylation; PRC1—Polycomb group (PcG) repressive complex 1

demethylation of the maternal pronucleus, facilitated by the exclusion of DNA methyltransferase 1 (DNMT1) from the nucleus (Howell et al. 2001), as well as by TET-mediated hydroxylation (Inoue and Zhang 2011). However, imprinting control regions (ICRs), oocyte, and sperm-contributed differentially methylated

regions (DMRs), as well as several families of repeats, such as class II intracisternal A-particles (IAPs) and L1Md_A elements, retain high levels of DNA methylation throughout the preimplantation development (Lane et al. 2003; Smith et al. 2012). Minimum levels of DNA methylation are reached at the blastocyst stage, followed by postimplantation gain of methylation to typical somatic levels (Smith et al. 2012).

Genome-wide reprogramming of histone modifications also occurs during the preimplantation development (Fig. 2.1). Immediately after fertilization, the paternal pronucleus is stripped of sperm-specific proteins called protamines and repackaged with maternally stored histone variant H3.3 that is usually associated with chromatin regions actively transcribed (Torres-Padilla et al. 2006). Interestingly, deposition of H3.3 into the paternal genome by the histone chaperone regulator A (HIRA) is an important event for the establishment of pericentric heterochromatin, which is required for proper chromosome segregation during the first mitosis (van der Heijden et al. 2005; Santenard et al. 2010). Only a few hours later, during the first DNA replication, the canonical histone H3 variants are incorporated for the first time into the paternal genome (Santenard et al. 2010). Histone H3.3 within the male pronucleus becomes trimethylated at lysine 27 (H3K27me3) and this repressive histone mark, together with H3K9me1 (monomethylation of lysine 9) retained in pericentromeric regions and residual DNA methylation, serves as a substrate for pericentric heterochromatin formation mediated by the Polycomb group (PcG) repressive complex 1 (PRC1) (Puschendorf et al. 2008). In the female pronucleus histone H3.3 transiently disappears and is replaced by histone H3.2 (Akiyama et al. 2011). After the two-cell stage, H3.1 and H3.3 variants re-localize to heterochromatin and euchromatin, respectively (Akiyama et al. 2011). The pericentric heterochromatin in the female pronucleus is marked with the repressive histone marks H3K9me3, H4K20me3, and H3K64me3 and binds HP1 β (heterochromatin 1 beta) protein (Santos et al. 2005; Probst et al. 2007; Daujat et al. 2009). Histone H2A variants are also reprogrammed during the preimplantation development. H2AZ (important for gene silencing), macroH2A (associated with heterochromatic regions and inactive X chromosome in females), and the canonical H2A are not incorporated into chromatin during the early cleavage stages, and are possibly even actively removed after fertilization. In contrast, H2AX (implicated in DNA repair) is particularly enriched in early embryos (Nashun et al. 2010). Together, all the reprogramming events described above are thought to contribute to the efficient acquisition of totipotency during preimplantation development.

The earliest sign of cell differentiation occurs at the blastocyst stage (embryonic day—E3.5 in mouse and embryonic day 5 in human), with the specification of the inner cell mass (ICM) and the trophectoderm (TE). This event coincides with the first wave of de novo DNA methylation. As result, TE is relatively hypomethylated compared with ICM, as revealed by the 5-methylcytosine staining (Santos et al. 2002). Similar to DNA methylation, several histone modifications, including H3K27 methylation, H3K9Ac (histone H3, lysine 9 acetylation), H4 acetylation, and H3K9 methylation, also exhibit asymmetry between ICM and TE, either at

global level or at specific loci (Erhardt et al. 2003; Sarmiento et al. 2004; O'Neill et al. 2006). The ICM is then separated during the late blastocyst stage into epiblast, that will form the future embryo proper, and primitive endoderm that contributes, together with the trophoblast cells derived from TE, to the formation of the extraembryonic tissues (Reik et al. 2001). Some cells derived from the primitive endoderm also contribute to the formation of the definitive embryonic endoderm (Kwon et al. 2008). After the formation of the three lineages (epiblast, primitive endoderm, and trophoblast), the cells undergo successive steps of differentiation to form all cell types of the organism, including placenta.

2.3 Epigenetic Regulation of Pluripotency

The three lineages at the blastocyst stage have been used for derivation of distinct stem cell types that can be maintained in vitro (Fig. 2.1): trophoblast stem (TS) cells from TE (Tanaka et al. 1998), extraembryonic endoderm stem (XEN) cells from the primitive endoderm (Kunath et al. 2005), and embryonic stem (ES) cells from the epiblast (Matsui et al. 1992). Analyses performed on these cell lines enabled the identification of key genetic factors that regulate pluripotency, such as OCT4 (octamer-binding transcription factor 4, also known as POU5F1—POU domain, class 5, transcription factor 1), NANOG (Nanog homeobox), SOX2 (SRY [sex-determining region Y]-box 2), and SALL4 (Sal-like protein 4) for ES cells (Mitsui et al. 2003; Loh et al. 2006; Wu et al. 2006); CDX2 (caudal type homeobox 2), EOMES (eomesodermin), and TEAD4 (TEA domain family member 4) for TS cells (Niwa et al. 2005; Yagi et al. 2007); and GATA4 (GATA-binding protein 4), GATA6 (GATA-binding protein 6), SOX7 (SRY [sex-determining region Y]-box 7), and SOX17 (SRY [sex-determining region Y]-box 17) for XEN (Kunath et al. 2005; Lim et al. 2008). The study of these stem cell lines also revealed an intriguing interplay between pluripotency transcription factors and epigenetic mechanisms. In fact, it is now thought that the dynamic balance between these two regulatory systems may form the basis for the pluripotent state.

2.3.1 DNA Methylation

The promoter regions of many pluripotency genes are unmethylated in pluripotent stem cell lines but methylated in somatic cells (Fouse et al. 2008; Meissner et al. 2008; Farthing et al. 2008; Senner et al. 2012). DNA methylation is thought to be particularly important for the epigenetic regulation of some “gatekeeper” genes that reinforce the commitment of pluripotent stem cells to a certain lineage such as *Elf5*, (E74-like factor 5 [ets domain transcription factor]) which, together with *Cdx2* and *Eomes*, safeguards ES cells from differentiating into trophoblast derivatives (Ng et al. 2008). Moreover, DNA methylation is the only epigenetic

mechanism that represses the activity of some genes implicated in differentiation of ES cells, and lack of DNA methylation in mutant ES cells leads to activation of these genes (Fouse et al. 2008). ES cells also contain substantial levels of non-CpG methylation (Ramsahoye et al. 2000; Lister et al. 2009; Ziller et al. 2011). Increased levels of non-CpG methylation have been found in exons of highly expressed genes, such as *OCT4* (Lister et al. 2009). Despite these characteristics, whether DNA methylation is absolutely necessary to maintain the pluripotency state remains controversial. Indeed, ES cells lacking completely DNA methylation (triple knock-out for *Dnmt1*, *Dnmt3a*, and *Dnmt3b*) can grow robustly and maintain to a large extent their undifferentiated characteristics (Tsumura et al. 2006).

The cytosines in DNA can acquire alternative modifications besides 5-methylcytosine. Tet1 protein, which is highly expressed in ES cells, can further modify 5-methylcytosine into 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine (Ito et al. 2011; Wu et al. 2011; He et al. 2011). High levels of 5-hydroxymethylcytosine and 5-formylcytosine in ES cells are associated with actively transcribed genes, as well as with Polycomb-repressed developmental regulators, and were demonstrated to guard against trans-differentiation to extra-embryonic lineages (Ficz et al. 2011; Wu et al. 2011; Booth et al. 2012; Raiber et al. 2012). However, deletion of *Tet1* gene in mice is compatible with embryonic and postnatal development, possibly due to partial compensation by *Tet2*. Accordingly, mutant ES cells display only subtle changes in gene expression and skewed differentiation towards trophectoderm in vitro (Dawlaty et al. 2011).

2.3.2 Histone Modifications

In addition to DNA modifications, histone modifications are also important in controlling gene expression during cell renewal of pluripotent stem cell lines. In agreement with the previously established notion that H3K4me3 is an activating histone modification (Santos-Rosa et al. 2002), peaks of this mark are observed in ES cells in association with promoter regions of key pluripotency genes (Azuara et al. 2006; Barski et al. 2007). However, approximately 2,000 genes that are transcriptionally repressed in ES cells but are required for later differentiation (such as *Sox*, *Hox*, *Fox*, *Pax*, and *Irxa* gene families) are concomitantly decorated at their promoter regions with both active H3K4me3 and repressive H3K27me3, pattern dubbed as “bivalent” (Azuara et al. 2006; Bernstein et al. 2006; Pan et al. 2007; Zhao et al. 2007). The bivalent domains are often found at promoters containing CpG islands and many bind OCT4, NANOG, or SOX2 (Bernstein et al. 2006; Mikkelsen et al. 2007). Virtually all bivalent domains bind PcG proteins belonging to the PRC2 complex (embryonic ectoderm development—EED, AE-binding protein 2—AEBP2, SUZ12—suppressor of zeste 12 homolog [*Drosophila*], and the H3K27 methyltransferase EZH2—enhancer of zeste homolog 2 [*Drosophila*]) (Ku et al. 2008). Recently, jumonji, AT-rich interactive domain 2 (JARID2—a member of the Jumonji family of lysine demethylases) and the

Polycomb-like 2 (PCL2) protein were also found to associate with PRC2 in mouse ES cells and are thought to play important roles in pluripotency (Peng et al. 2009; Pasini et al. 2010; Walker et al. 2010). A subset of the bivalent domains also binds PcG proteins of the PRC1 complex (Ku et al. 2008). The PRC1 protein RNF2 (ring finger protein 2, also known as RING1B) is responsible for ubiquitination of histone H2A at lysine 119, which in turn is responsible for RNA polymerase II stalling (a mechanism for transcriptional silencing) at promoters of bivalent genes (Stock et al. 2007). The H3K4me3 mark at the bivalent domains is generated by the H3K4 methyltransferase activity of the MLL/trithorax complex (myeloid/lymphoid or mixed-lineage leukemia [trithorax homolog, *Drosophila*]) (Dou et al. 2006). Similar bivalent chromatin profiles have also been identified in TS cells at promoter regions of silenced, lineage-specific regulatory genes (Santos et al. 2010). However, in XEN cells lineage-specific genes are marked solely by repressive histone modifications, pattern thought to reflect the restricted developmental potential of these cells (Santos et al. 2010).

The repressive histone modifications H3K9me2 and H3K9me3 colonize different regions of the ES cells' genome. H3K9me2 is found in large blocks in the genome (several megabases each), which are highly conserved between mouse and human (Wen et al. 2009). H3K9me3, found in ES cells mostly in partnership with H3K20me3, is required for silencing several classes of endogenous retroviruses (Mikkelsen et al. 2007; Rowe et al. 2010; Matsui et al. 2010; Macfarlan et al. 2011).

In addition to promoters and repetitive DNA, histone modifications are particularly important in regulating the activity of enhancer elements (short regions of DNA often found distant from transcription start sites that bound transcription factors and enhance gene transcription). Based on the patterns of histone modifications, two distinct classes of enhancers can be identified in ES cells. Both classes are characterized by open chromatin, marked by the presence of DNase I hypersensitive sites (DHSs), enrichment in highly mobile nucleosomes containing the specialized histone variants H3.3 and H2A.Z, binding of the histone acetyltransferase P300, and monomethylation of histone H3 at lysine 4 (H3K4me1). Active enhancers, often located in the vicinity of active genes such as the key pluripotency genes, are characterized by acetylation of histone H3 at lysine 27 (H3K27ac). In contrast, the so-called poised enhancers, located near genes involved in controlling early steps of differentiation and marked with bivalent domains at their promoters, are depleted in H3K27ac and instead are enriched in H3K27me3 and H3K9me3 (Creyghton et al. 2010; Rada-Iglesias et al. 2011; Zentner et al. 2011; Buecker and Wysocka 2012).

2.3.3 Chromatin-Modifying Complexes

With the exception of acetylation, most histone modifications do not impose directly changes in the chromatin conformation. Instead, they often bind chromatin-remodeling factors, which utilize the energy released from ATP

hydrolysis to exchange histones and reposition or evict nucleosomes. When compared with differentiated cells, pluripotent stem cells are characterized by a generally open chromatin state (Gaspar-Maia et al. 2011). There are four families of chromatin remodelers (SWI/SNF, CHD/NURD, ISWI, and INO80) and many subunits of these families have been shown to play important roles in pluripotent stem cells (reviewed by Gaspar-Maia et al. 2011). For example, the SWI/SNF family member BRG1 (also known as SMARCA4) opposes PcG proteins by opening the chromatin at LIF/STAT3 (leukemia inhibitory factor/signal transducer and activator of transcription 3) target sites in ES cells. However, BRG1 also facilitates PcG function at classical PcG targets, including all four *Hox* loci, reinforcing their repression in ES cells (Ho et al. 2011). The chromodomain helicase DNA-binding protein 1 (CHD1) member of the CHD family binds globally to active euchromatin and co-localizes with RNA polymerase II (RNAPII) in ES cells and CHD1 depletion by RNA interference leads to accumulation of high levels of heterochromatin (Gaspar-Maia et al. 2009). CHD3 and CHD4 constitute the catalytic subunit of the nucleosome-remodeling (NuRD) complex, which also contains histone deacetylases (HDAC1 and HDAC2) and a methyl-binding protein (MBD3). MBD3 cooperates with BRG1 to maintain the global levels of 5-hydroxymethylcytosine in ES cells (Yildirim et al. 2011). Finally, the TIP60/KAT5–P400 (lysine acetyltransferase 5/E1A-binding protein p400) complex of the INO80 family facilitates transcription by combining nucleosome remodeling with histone acetylase activity. ES cells depleted in different subunits of the TIP60–P400 complex exhibit decreased proliferation rates, reduced pluripotency, and reduced viability (Fazzio et al. 2008). The TIP60–P400 complex also binds H3K4me3 at bivalent domains, an interaction that is facilitated by Nanog (Fazzio et al. 2008).

2.3.4 Mechanisms for Targeting Epigenetic Patterns in Pluripotent Stem Cells

The patterns of chromatin and DNA modifications in pluripotent stem cells cannot be explained only by the genomic distribution of transcription factor-binding sites. Indeed, targeting of PcG proteins at the bivalent promoters in ES cells is only partially explained by the concomitant binding of the core pluripotency transcription factors (Bernstein et al. 2006; Mikkelsen et al. 2007). Therefore, other factors such as the local DNA sequence, noncoding RNAs (ncRNAs), and the higher order chromatin structure may be important in this process.

The direct involvement of DNA sequence was first demonstrated in the case of JARID2, protein that binds directly to DNA and plays a major role in targeting PRC2 complexes to the correct sites (Peng et al. 2009; Pasini et al. 2010). Additionally, a recent study has identified a PcG responsive element in human ES cells, a highly conserved 1.8 kb DNA sequence located between *HOXD11* (homeobox D11) and *HOXD12* (homeobox D11) genes, which is nucleosome depleted and GC-rich and contains YY1 transcription factor-binding sites (Woo et al. 2010).

It has also been demonstrated that short DNA sequences inserted into the mouse ES cells can autonomously induce hypo- or de novo methylation in *cis* (Lienert et al. 2011).

ncRNAs may be another important class of regulators for establishing epigenetic patterns in ES cells. Short ncRNAs (<200 nt) interact with PRC2 and are involved in stabilizing PRC2 association with chromatin, though the importance of direct base pairing at specific sequence motifs is still unknown (Kanhare et al. 2010). Furthermore, over 3,000 large intergenic noncoding RNAs (lincRNAs) have been recently identified in mouse ES cells. At least a third of them are associated with chromatin complexes involved in reading, writing, or erasing histone modifications and are critical for pluripotency maintenance (Guttman et al. 2009, 2011).

Finally, the higher order chromatin structure may also contribute to the appropriate establishment of epigenetic marks in pluripotent stem cells (reviewed by Li et al. 2012). The insulator protein CTCF (CCCTC-binding factor [zinc finger protein]), known to mediate long-range interactions between distant regulatory elements, has been shown to cooperate with Oct4 in organizing the chromatin loops at the *Nanog* locus (Levasseur et al. 2008). A recent genomic analysis in ES cells has revealed that binding sites for cohesin (a key partner for CTCF), mediator, and the cohesin-loading factor NIPBL (nipped-B homolog [*Drosophila*]) overlap at active promoters and enhancers (Kagey et al. 2010). Together, these complexes contribute to chromatin looping between enhancers and promoters in patterns that are specific to ES cells (Kagey et al. 2010).

2.4 Establishment of Epigenetic Patterns During Differentiation

Gene deletion studies in mice or knockdown experiments in ES cells demonstrated for the first time that many epigenetic modifiers play critical roles during differentiation. Accordingly, deletion of *Dnmt1* results in lethality before E10.5 (Li et al. 1992) and disruption of *Dnmt3b* is lethal before E9.5 (Okano et al. 1999). Additionally, ES cells deficient for all three DNA methyltransferases show increased cell death upon differentiation into epiblast lineages, but not during differentiation into extraembryonic lineages and do not contribute to embryonic lineages when injected into blastocysts (Sakaue et al. 2010). Deletion of the histone methyltransferase G9a/EHMT2 (euchromatic histone-lysine N-methyltransferase 2) results in embryonic lethality between E8.5 and E9.5 (Tachibana et al. 2002). Knockdown of various PcG proteins in ES cells affects their ability to differentiate (Azuara et al. 2006; Bernstein et al. 2006; Pasini et al. 2007, 2010). Depletion of the MLL complex component DPY30 (dpy-30 homolog [*C. elegans*]) in ES cells, which decreases H3K4me3 at bivalent domains, results in a significant reduction in the differentiation potential, particularly along the neural lineage (Jiang et al. 2011). Deletions of *Mbd3* or *Hdac1* result in aberrant differentiation of mouse ES cells (Kaji et al. 2006; Dovey et al. 2010).

In addition to these loss-of-function studies, another major approach for improving our understanding of the transitions that occur during differentiation was the use of various “omics” analyses. Most studies that addressed the role of epigenetic modifications during differentiation have compared the genomic distribution of various marks between ES cells and cells differentiated in vitro or with donor-derived somatic cells. These studies have identified several general mechanisms implicated in the establishment of tissue-specific epigenetic patterns during differentiation.

2.4.1 Dynamics of DNA Methylation During Differentiation

Whereas a small number of genes undergo DNA demethylation upon commitment to a cell lineage, many more gain CpG methylation (Fouse et al. 2008; Meissner et al. 2008). Loss of DNA methylation (Fig. 2.2a) is observed especially at lineage-specific gene-regulatory elements. De novo DNA methylation (Fig. 2.2b) is responsible for the active repression of core pluripotency and germline-specific genes, as well as for some lineage choice events. Repression of pluripotency genes is initiated by local binding of the G9a histone methyltransferase, which through its SET domain brings about local methylation of histone H3K9me3 (Feldman et al. 2006; Epsztejn-Litman et al. 2008). Subsequently, H3K9me3 binds HP1/CBX5 (chromobox homolog 5), thus generating a local heterochromatic structure. In parallel, G9a/EHMT2 recruits the DNA methyltransferases Dnmt3a and 3b, which then induce de novo methylation of these genes (Feldman et al. 2006; Epsztejn-Litman et al. 2008). De novo DNA methylation during ES cell differentiation also occurs at CpG island promoters and at sequences outside of promoter regions, many of which act as enhancers (Mohn et al. 2008; Meissner et al. 2008; Stadler et al. 2011). Importantly, a recent study performed on early embryos confirmed that Dnmt3b catalyzes the gain of DNA methylation in E6.5 epiblast cells (Borgel et al. 2010). Similarly with the data obtained in ES cells cultured in vitro, these epigenetic events target promoters of pluripotency and germline-specific genes, as well as genes programmed to be expressed later during development. For the latter category of genes, promoter methylation acquired in epiblast is then erased during terminal cell differentiation (Borgel et al. 2010).

2.4.2 Resolution of Bivalent Domains

In the case of bivalent domains (Fig. 2.2c) it is thought that the concomitant presence of active and repressive modifications in pluripotent stem cells allows rapid resolution of these domains into single H3K27me3 or H3K4me3 marks during differentiation (Mikkelsen et al. 2007). Removal of H3K27me3 is achieved by two H3K27 demethylases: UTX/KDM6A (lysine [K]-specific demethylase 6A) and JMJD3/KDM6B (lysine [K]-specific demethylase 6B) (Agger et al. 2007; Lee et al. 2007; Lan et al. 2007). For example, human ES cells induced to differentiate

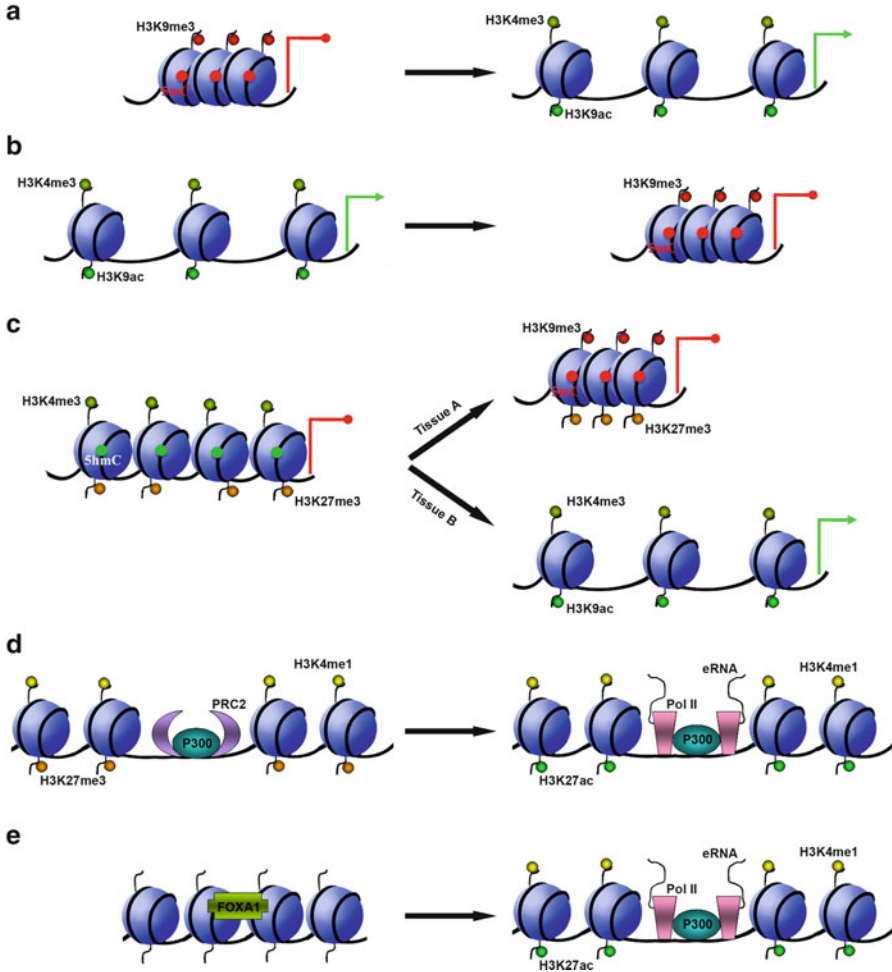


Fig. 2.2 Epigenetic changes during differentiation (see text for details). (a) Loss of DNA methylation at lineage-specific regulatory genes (together with changes in histone marks) leads to gene activation. (b) De novo DNA methylation at promoters of pluripotency genes or germline-specific genes. (c) Epigenetic reprogramming of bivalent promoters allows gene silencing or activation in a tissue-specific manner. (d) Transition between the “poised” and the “active” state at enhancer elements during differentiation. (e) The role of “pioneer” transcription factors (such as FoxA1—forkhead box A1) in generating de novo enhancers during differentiation. K—lysine; me—methylation; ac—acetylation; P300—histone acetyltransferase; PRC2—Polycomb group (PcG) repressive complex 2; Pol II—RNA polymerase II; eRNA—enhancer-associated RNAs

by treatment with retinoic acid recruit KDM6A at the promoters of the anterior genes of *HOXA* and *HOXB* loci. Recruitment of KDM6A to these promoters coincides with disappearance of H3K27me3, decreased occupancy of the PRC2 complex components SUZ12 and EZH2, and gene activation, while knockdown of

KDM6A prevents these events (Agger et al. 2007; Lee et al. 2007). KDM6A has also been demonstrated to activate muscle-specific genes during myogenesis, being targeted to the correct promoters by the transcriptional activator Six4 (SIX homeobox 4) (Seenundun et al. 2010). Interestingly, KDM6A associates with two H3K4 methyltransferases, MLL3/KMT2C (myeloid/lymphoid or mixed-lineage leukemia 3) and MLL4/KMT2D (myeloid/lymphoid or mixed-lineage leukemia 4), suggesting cooperation between H3K4 methylation and H3K27 demethylation (Lee et al. 2007; Issaeva et al. 2007). JMJD3/KDM6B has been demonstrated to resolve the bivalent domain at the *Nes* (nestin) gene promoter and to control the expression of key regulators and markers of neurogenesis during the commitment of ES cells towards the neural lineage (Burgold et al. 2008). Removal of H3K4me3 from the bivalent domains is achieved by the KDM5 demethylases. KDM5A (JARID1A/RBP2) is recruited at the bivalent domains by the PRC2 complex (Pasini et al. 2008). Additionally, KDM5B (JARID1B/PLU1) binds to a substantial fraction of bivalent domains in ES cells and is required for silencing stem cell and germ cell-specific genes during ES cell differentiation into neural progenitor cells (Schmitz et al. 2011).

It is important to stress that bivalent domains play important roles throughout differentiation. Indeed, when ES cells are differentiated into neural cells, the resolution of some bivalent domains is counterbalanced by appearance of new ones at other promoter regions. Moreover, ~41 % of the bivalent domains found in ES cells are preserved after differentiation into terminal pyramidal neurons (Mohn et al. 2008). In hemangioblasts, which are hematopoietic/endothelial precursors, some neuronal genes retain bivalency and require the presence of the PRC1 component RING1B/RNF2 to remain silent (Mazzarella et al. 2011). Adult stem cells, which maintain the natural homeostasis of adult tissues by supplying a continuous pool of differentiated cells in response to external signals, also contain bivalent chromatin domains (Mikkelsen et al. 2007; Cui et al. 2009).

2.4.3 *Chromatin Changes at Enhancer Elements*

Epigenetic reprogramming at enhancer elements is perhaps one of the most important events in the establishment of tissue-specific gene expression patterns during differentiation. Indeed, during differentiation of human ES cells into a mesendodermal lineage, chromatin modifications at promoters remained largely invariant, with much greater dynamics in chromatin modifications at enhancers, especially for H3K4me1 and H3K27ac (Hawkins et al. 2011). The main event that takes place during differentiation at enhancers is a switch from the poised to the active status (Fig. 2.2d), which coincides with the ability to drive gene expression. Interestingly, recent evidence suggests that the pluripotency factors active in ES cells are not only involved in maintaining the pluripotent state of these cells but also have a direct role during differentiation. Accordingly, SOX2, which binds many poised enhancers in ES cells, is replaced by SOX3 in neural progenitor cells and then by SOX11 in differentiated neurons. Upon binding of activating SOX3 or

SOX11 transcription factors, the poised chromatin state is resolved into an active one (Bergsland et al. 2011). Other poised enhancers are co-occupied by several pluripotency factors in ES cells and this multiple binding is thought to prevent their premature activation. One example is the *EOMES* enhancer, which is bound by OCT4, SOX2, and NANOG in human ES cells (Teo et al. 2011). At the onset of endoderm specification, SOX2 departure and the persistence of NANOG binding lead to activation of this enhancer and increased expression of eomesodermin (Teo et al. 2011).

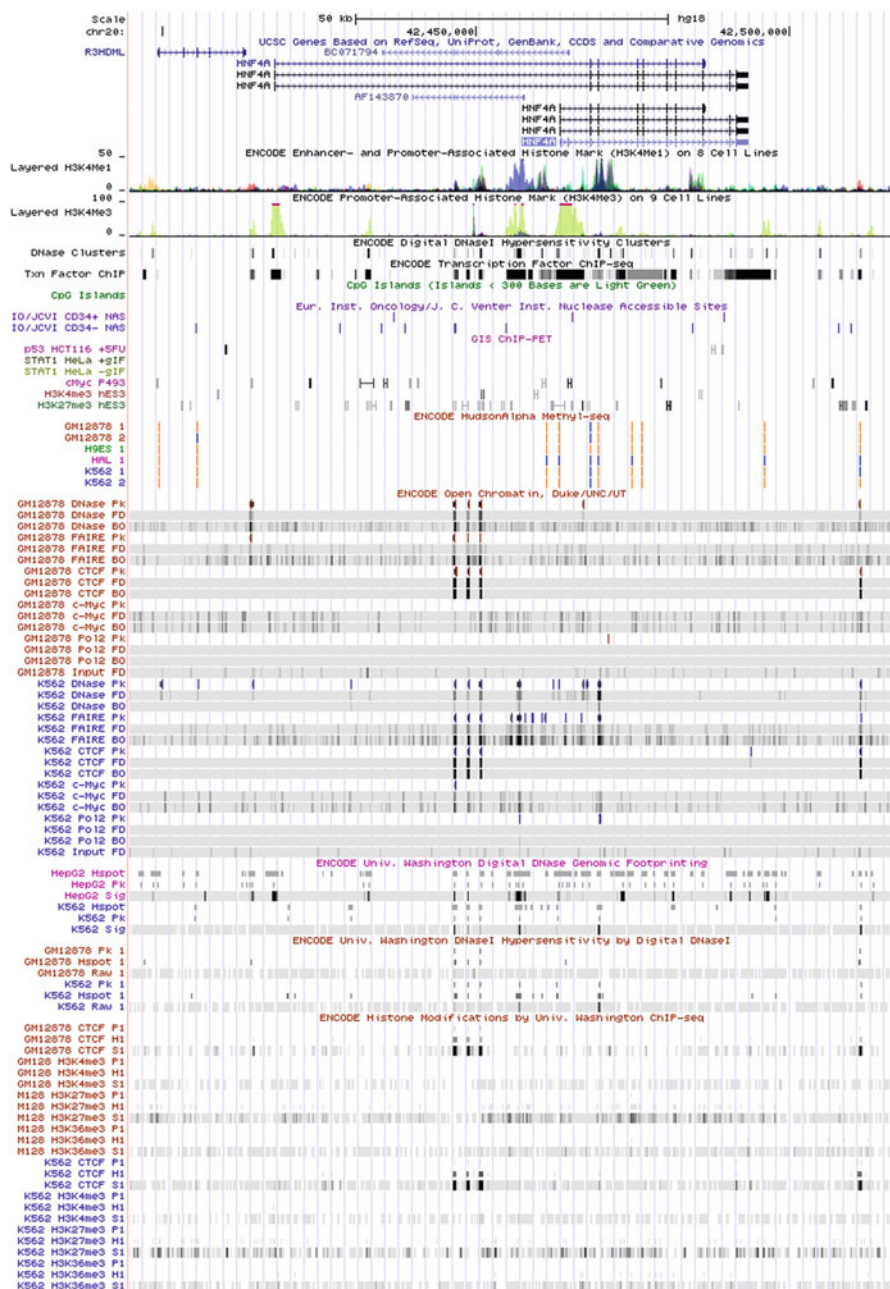
Not all enhancers that are used during differentiation are established in pluripotent cells as active or poised ones. Many enhancers are generated de novo during differentiation by the intervention of the so-called pioneer factors, which are often lineage-specific transcription factors that have the ability to bind DNA sequences at chromatin-compacted regions (Fig. 2.2e). Pioneer factors such as FOXA (forkhead box A) and GATA (GATA-binding protein) recruit subsequently chromatin remodelers, which establish the characteristic open chromatin structure of active enhancers (Zaret and Carroll 2011).

2.5 Cell Type-Specific Epigenetic Patterns in Differentiated Mammalian Cells: Lessons Learned from the ENCODE Project

A first indication on the complexity of cell type-specific epigenetic patterns came from the analysis of just 1 % of the human genome during the pilot phase of the ENCODE project (ENCODE Project Consortium 2007). This preliminary analysis was then extended to the entire genome during the production stage of the ENCODE project. This stage of the project ended up with the release of a much more extensive set of results (Fig. 2.3) obtained in 147 different cell types (including both immortalized cell lines and primary cell types from a variety of tissues and developmental stages). The most important conclusion of the project was that approximately 80 % of the human genome participates in at least one biochemical function, most of which are related to gene regulation (ENCODE Project Consortium 2012). Some of the most relevant findings of the project are summarized below.

2.5.1 *Transcriptional Landscape*

RNA sequencing was performed in the set of 15 cell types most commonly used across the consortium and showed that, although the 20,687 protein-coding genes cover less than 3 % of the genome, cumulatively, nearly 75 % of the human genome is transcribed (Djebali et al. 2012). Each protein-coding gene associates on



average 6.3 alternatively transcripts and, although many isoforms are expressed simultaneously in a single cell type, one of these dominates (Djebali et al. 2012). The majority of protein-coding genes (53 %) are constitutively expressed (in all cell lines) and only a small fraction (7 %) are cell line specific (Djebali et al. 2012). ENCODE also identified 9,277 manually curated long noncoding RNA (lncRNA) loci generating ~15,000 transcripts, most of which are associated with chromatin and display more tissue-specific expression patterns than the protein-coding genes (Derrien et al. 2012). Approximately 18 % of the protein-coding and lncRNA genes exhibit allele-specific expression (Djebali et al. 2012). ENCODE also identified 11,216 pseudogenes, of which 876 are transcribed (Pei et al. 2012). Only a small fraction of the transcribed pseudogenes are active in all tissues analyzed, while most are transcribed only in one tissue (Pei et al. 2012). There are also 7,053 annotated small RNAs, which include 1,944 small nuclear (sn)RNAs, 1,521 small nucleolar (sno)RNAs, 1,756 μ (mi)RNAs, and 624 transfer (t)RNAs (Djebali et al. 2012). Other categories of transcripts include unannotated short RNAs such as the promoter-associated short RNAs (PASRs) and the terminus-associated short RNAs (TASRs), transcripts emanating from repeat elements and enhancer-associated RNAs (eRNAs) (Djebali et al. 2012).

2.5.2 DNA Methylation Landscape

DNA methylation was analyzed in 82 cell types (cell lines and primary cells) using the reduced representation bisulfite sequencing (RRBS), which can interrogate 1.2 million CpGs located in intergenic regions, proximal promoters, and gene bodies (8.6 % of non-repetitive genomic CpGs), with a preferential bias towards CpG islands (Meissner et al. 2008). Ninety-six percentage of all analyzed CpGs were found to exhibit differential methylation in at least one cell type, with the highest variability found at gene bodies and intergenic regions, rather than at promoters (ENCODE Project Consortium 2012). In addition, unmethylated intragenic CpG islands were found to associate binding of P300 histone acetyltransferase, a known marker for enhancer activity (Creyghton et al. 2010; ENCODE Project Consortium 2012). Differential DNA methylation associates with tissue-specific binding of CTCF, a ubiquitously expressed regulator of transcription and chromatin structure. Comparison between DNA methylation distribution and CTCF-binding sites for a subset of ~4,000 CTCF peaks indicates that over 40 % of the cell type-specific CTCF binding is associated with local differential DNA methylation (Wang et al. 2012). Additionally, 20 % of the DHSs with cell type-specific accessibility show a significant negative correlation with levels of DNA methylation, while the remaining 80 % of DHSs are constitutively hypomethylated (Thurman et al. 2012; Neph et al. 2012). Moreover, for 70 % of transcription factors, average methylation at cognate binding sites is significantly and negatively correlated with transcript levels of the corresponding transcription factors (Thurman et al. 2012).

2.5.3 *Histone Modification Landscape*

The ENCODE project analyzed systematically 11 histone modifications and 1 histone variant (H2A.Z) in 46 cell types. The main conclusion of this analysis is that histone modification patterns can be reliably used to assign functional attributes to genomic regions (ENCODE Project Consortium 2012; Dong et al. 2012). For example, transcriptionally active GC-rich (and TATA-less) promoters are associated with H2A.Z, H3K9ac, H3K27ac, H3K4me3, and H3K4me2, while repressed promoters are associated with H3K27me3 or H3K9me3. H3K79me2 and H3K36me3 are marks of transcription elongation; however, H3K79me2 occurs preferentially at the 5' end of the gene bodies, while H3K36me3 is enriched at 3' of the first intron (ENCODE Project Consortium 2012). These two last marks can also be used to predict patterns of alternative splicing: H3K36me3 has a positive contribution to exon inclusion, while H3K79me2 has a negative contribution (ENCODE Project Consortium 2012). By overlapping histone patterns with DHS maps 44,853 novel putative promoters were identified, many of which are active in a cell-specific manner, are contained within the gene bodies of previously annotated genes, and show antisense orientation (Thurman et al. 2012). Patterns of histone modifications at enhancer regions are amongst the best associated with cell-specific gene activity. Active enhancers are characterized by the presence of DHSs that bind RNA polymerase II and are enriched in H3K4me1, H3K27ac, H3K9ac, and H3K79me2 and depleted in H3K27me3 (Thurman et al. 2012; Djebali et al. 2012).

2.5.4 *Open Chromatin Landscape*

Regions of open chromatin identified by DNase I hypersensitivity are often found at regulatory DNA regions. Using DNase-seq in 125 cell types ~2.9 million DHSs were identified, most of them being located distal to transcription start sites (TSSs) and highly cell specific. A complementary technique—FAIRE-seq (formaldehyde-assisted isolation of regulatory elements) performed in 25 cell types—also identified ~4.8 million sites depleted in nucleosomes, many of which overlap with DHSs (ENCODE Project Consortium 2012; Thurman et al. 2012). Overlapping DHSs with high-throughput ChIP-seq data for 42 transcription factors in the K562 cell line (immortalized myeloid leukemia cells) showed that over 94 % of the transcription factor-binding sites fall within accessible chromatin. Notable exceptions are transcription factors known to bind to compacted heterochromatin, such as TRIM28 (tripartite motif containing 28), SETDB1 (SET domain, bifurcated 1), and ZNF274 (zinc finger protein 274) (Thurman et al. 2012). Moreover, a correlation between distal DHSs and DHSs located at known promoters across 79 cell types allowed functional connection of ~580,000 distal enhancers with their target promoters. Most promoters are connected with more than one distal DHS and vice versa, indicating a very complex *cis*-regulatory circuit of the human genome. In addition to this synchronized activation between promoters and distal enhancers,

hundreds of enhancers around the genome showed patterns of matched co-activation, suggesting highly choreographed cell type-specific behavior and common functions (Thurman et al. 2012). Finally, micrococcal nuclease (MNase) digestion followed by high-depth sequencing was used to map nucleosome occupancy in two cell types: GM12878 (lymphoblastoid cell line) and K562 (Kundaje et al. 2012). This analysis, combined with 12 histone marks, DNase-seq, and binding sites for 119 DNA-binding proteins, demonstrated that, with the exception of CTCF/cohesion complex, nucleosomes as well as histone marks are deposited asymmetrically around promoters, enhancers, or transcription factor-binding sites (Kundaje et al. 2012).

2.5.5 Long-Range Interaction Landscape

Physical interactions between distant chromosomal regions, which are thought to be important for regulation of gene expression, were assessed using two complementary technologies: 5C (chromosome conformation capture carbon copy) and ChIA-PET (chromatin interaction analysis with paired-end tag sequencing) (ENCODE Project Consortium 2012; Sanyal et al. 2012). The 5C approach was used for an unbiased interrogation of all interactions between TSSs previously identified by the pilot ENCODE project and distal genomic regions. This assay performed in four cell types identified over 1,000 long-range interactions (Sanyal et al. 2012). The most frequent interactions of the assessed TSSs were with enhancers, other promoters, and CTCF-binding sites, and each of these elements was found to be engaged in multiple interactions. The TSS–enhancer and TSS–promoter interactions were often found to be cell type specific, while the interactions of TSS–CTCF were most of the time common to all four cell types (Sanyal et al. 2012). The ChIA-PET approach, which interrogates interactions between chromatin regions that bind RNA polymerase II, has been applied within the ENCODE project for the K562 cell line. This analysis identified over 120,000 promoter-centered interactions, the vast majority of which were intrachromosomal. Similar to the 5C approach, this analysis showed that most promoters are engaged in multiple promoter–enhancer and promoter–promoter interactions (ENCODE Project Consortium 2012). The ChIA-PET has also been used in an independent study performed in five human cell types (including K562) (Li et al. 2012). This study demonstrated widespread promoter–promoter interactions between genes transcribed cooperatively, as well as cell type-specific promoter–enhancer interactions (Li et al. 2012).

2.6 Tissue-Specific Epigenetic States and Human Disease

It is increasingly acknowledged that epigenetic phenomena may be a crucial component in the development of human disease. The importance of epigenetics has been clearly demonstrated in monogenic disorders involving imprinted genes

(such as Beckwith–Wiedemann, Prader–Willi, and Angelman syndromes), in single-gene disorders of the epigenetic machinery (such as Rett, ICF, ATRX, and Rubinstein–Taybi syndromes) and in cancer (reviewed by Feinberg 2007; Portela and Esteller 2010). Since these subjects are being presented in depth elsewhere in the book, in this section I discuss the existing evidence for tissue-specific epigenetic alterations in common diseases and the link between genetic variants, tissue-specific epigenetic patterns, and disease.

2.6.1 Epigenetic Alterations in Common Human Diseases

The tissue specificity of epigenetic patterns makes it less straightforward to extrapolate the epigenetic information obtained in accessible samples such as peripheral white blood or buccal cells to the relevant tissues involved in various human diseases. Additional obstacles in studying epigenetic alterations in the context of human diseases are the observed variation of epigenetic marks between healthy individuals and with advancing age (Sandovici et al. 2003; Bjornsson et al. 2008). Despite these important challenges, in the past decade a number of studies have been able to uncover epigenetic alterations in several major forms of common human diseases in tissues and at loci that are directly involved in the pathogenesis of the studied conditions.

By far, the most studied epigenetic mark until now was DNA methylation, fact explained at least in part by the ease of obtaining DNA samples compared with good-quality chromatin, as well as by the robustness and relatively low cost of new microarray-based technologies. For example, a study performed in monozygous twins (MZ) discordant for type 1 diabetes (T1D) using Illumina Infinium 27 K microarrays to measure DNA methylation in CD14⁺ monocytes identified 58 CpG sites hypermethylated and 78 hypomethylated in the T1D-affected co-twins (Rakyan et al. 2011). Using the same technology to measure DNA methylation in the CD4⁺ T lymphocytes from patients with systemic lupus erythematosus and controls 236 hypomethylated and 105 hypermethylated CpG sites were identified (Jeffries et al. 2011). Another study using Illumina Infinium 27 K microarrays identified 276 CpG loci affiliated to promoters of 254 genes displaying significant differential DNA methylation in islets from type 2 diabetes (T2D) patients compared with controls, 244 of which were hypomethylated. These methylation changes affected many genes implicated in β -cell survival and function, were absent in blood cells from T2D individuals, and could not be induced experimentally in nondiabetic islets exposed to high glucose (Volkmar et al. 2012). As a last example, a study performed in the frontal cortex of patients with schizophrenia or bipolar disorder versus controls using CpG-island microarrays identified DNA methylation differences at dozens of loci, including several involved in glutamatergic and GABAergic neurotransmission, brain development, and other processes functionally linked to these diseases (Mill et al. 2008).

A variety of histone marks associated with transcriptional activation or repression have been studied in several common diseases in a tissue-specific context.

For example, in human prefrontal cortexes from patients with schizophrenia, the decreased *GAD1* (glutamate decarboxylase 1 [brain, 67 kDa]) expression compared to controls was associated with decreased levels of promoter H3K4me3, especially in females (Huang et al. 2007). Altered levels of the repressive histone mark H3K9me2 have been found at many loci implicated in autoimmunity and inflammation in lymphocytes collected from T1D patients versus controls (Miao et al. 2008). Marked differences in H3K9Ac levels were found at the upstream regions of *HLA-DRB1* (major histocompatibility complex, class II, DR beta 1) and *HLA-DQB1* (major histocompatibility complex, class II, DQ beta 1) genes in T1D monocytes relative to controls (Miao et al. 2012). Differential distribution of H3K4me3 and H3K9me3 peaks across the genome has been identified in cardiomyocytes collected from patients with heart failure caused by dilated cardiomyopathy compared to controls and many disease-dependent clusters contained genes implicated in signal transduction pathways for cardiac function (Kaneda et al. 2009).

A number of recent studies have also started to identify important roles for tissue-specific epigenators in the pathogenesis of several common human diseases. One illustrative example was recently published in human pancreatic islets. A comprehensive strand-specific transcriptome analysis identified 1,128 lncRNAs, many of which are cell specific and linked with β -cell differentiation and maturation programs. Using a gene candidate approach, several of these genes were found abnormally expressed in samples collected from T2D patients (Morán et al. 2012).

2.6.2 Genetic Variants, Tissue-Specific Epigenetic Patterns, and Human Disease

The convergence between disease-associated genetic variants emerging from GWAS and epigenetic maps led to the remarkable observation that many single-nucleotide polymorphisms (SNPs) linked with various diseases are located at regulatory DNA sequences. For example, rs7903146, a *TCF7L2* (transcription factor 7-like 2 [T-cell specific, HMG-box]) intronic variant strongly associated with T2D was found to be located in an islet-selective open chromatin region that exhibits enhancer activity. Human islets heterozygous for rs7903146 showed allelic imbalance in the local chromatin organization and altered enhancer activity (Gaulton et al. 2010). Additionally, a systematic analysis of chromatin-state dynamics in several human cell types, which identified cell type-specific enhancers, found that top-scoring disease-associated SNPs are frequently positioned within enhancer regions specifically active in the relevant cell types. Accordingly, SNPs associated with erythrocyte phenotypes are located in enhancers specific to erythroleukemia cells (K562), SNPs associated with systemic lupus erythematosus are located in enhancers specific to lymphoblastoid cells (GM12878), while SNPs

associated with triglyceride and total lipid levels in blood are located in enhancers specific to hepatocellular carcinoma cells (HepG2) (Ernst et al. 2011).

Building on these initial observations, the recent data published by the ENCODE consortium demonstrated unequivocally that over a third of the disease-associated genetic variants emerged from the GWASs performed so far localize within regulatory DNA elements marked by the presence of DHSs (ENCODE Project Consortium 2012; Maurano et al. 2012; Schaub et al. 2012). Beyond this statistically significant concentration of disease-associated genetic variants around regulatory DNA elements, the systematic analysis of a large number of cell and tissue types led to several additional striking observations. First of all, it was observed that genetic variants associated with a certain disease are particularly enriched at DHSs that are active in the cell types implicated in its pathogeny. Examples include the enrichment of genetic variants associated with Crohn's disease at DHSs active in T cells (subtypes T_H17 and T_H1) and the enrichment of SNPs associated with multiple sclerosis at DHSs active in $CD3^+$ T cells from cord blood and $CD19^+/CD20^+$ B cells (Maurano et al. 2012). In many cases common SNPs associated with specific diseases are located at binding sites for key transcription factors and, as a result, their presence induces allelic imbalances of chromatin states (Gaulton et al. 2010; Maurano et al. 2012). Additionally, the disease-associated SNPs can also alter tissue-specific enhancer–promoter interactions (Li et al. 2012; Maurano et al. 2012). More than 80 % of DHSs containing disease-associated SNPs are active in fetal cells and tissues, with the greatest enrichment for SNPs linked with phenotypes for which gestation or early growth have been shown to play major roles (such as cardiovascular disease) and with a relative depletion for SNPs linked with aging-related diseases (Maurano et al. 2012). This finding is in agreement with the recurring theme of chromatin landscape plasticity during early development and the risk for specific common adult diseases (Sandovici et al. 2008, 2011).

2.7 Concluding Remarks and Future Perspectives

The recent epigenomics studies have begun to uncover in great details the epigenetic landscape of pluripotent stem cells and the transitions that occur during cell differentiation. Despite the advances made by these studies, our understanding of the functional role of particular epigenetic mechanisms remains relatively poor. This limitation highlights the need for more mechanistic studies. It became apparent that some of the epigenetic features discovered by studying pluripotent stem cells are acquired during their culture *in vitro* and may not reflect the patterns existing *in vivo*. For example, ES cells cultured in defined medium with inhibitors of two kinases (MEK [MAP kinase/ERK kinase] and GSK3B [glycogen synthase kinase 3 beta]), a condition known as “2i,” postulated to establish a naive ground state, have reduced prevalence of bivalent domains, despite similar differentiation potential with serum-grown ES cells (Marks et al. 2012). Additionally, genome-wide DNA methylation profiling of a large collection of human pluripotent stem

cell lines has revealed many aberrations of this epigenetic mark, which were specific to the culture conditions used (Nazor et al. 2012). Currently it is still difficult to investigate lineage specification and the associated establishment of global epigenetic patterns for many cell types without expanding them in vitro. However, as the number of cells required for epigenetic analyses continues to decrease, it is likely that these exciting studies will become possible in the near future.

The recent completion of the ENCODE project represents a milestone achievement for our better understanding of the human genome and epigenome. However, the information obtained is still not comprehensive. For example, just 11 of more than 60 known histone modifications were analyzed in only 46 out of 147 cell types included in ENCODE and the real number of histone modifications may be even larger (Tan et al. 2011). Most of the other assays were performed only in small subsets of cell types, suggesting that the data obtained so far may represent only a fraction of the potential functional information encoded in the human genome. An important future goal is therefore to complete these gaps, for example by current complementing international projects such as the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010), Alliance for the Human Epigenome and Disease (AHEAD) (Jones et al. 2008), and BLUEPRINT (Adams et al. 2012). Important continuations of the original ENCODE project are also provided by the modENCODE project, set to identify functional elements in selected model organisms (*Drosophila melanogaster* and *Caenorhabditis elegans*) (Celniker et al. 2009) and Mouse ENCODE project (Mouse ENCODE Consortium 2012). A better integration between ENCODE and GWAS data is likely to have a significant impact on our understanding of common human diseases. This will be further enhanced by the recent completion of the 1000 Genomes project (The 1000 Genomes Project Consortium 2012).

Acknowledgements I thank Gabriella Ficiz and Wendy Cooper for critical reading and helpful suggestions.

References

- Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S et al (2012) BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* 30:224–226
- Agger K, Cloos PA, Christensen J, Pasini D, Rose S et al (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 449:731–734
- Akiyama T, Suzuki O, Matsuda J, Aoki F (2011) Dynamic replacement of histone H3 variants reprograms epigenetic marks in early mouse embryos. *PLoS Genet* 7:e1002279
- Albert M, Peters AH (2009) Genetic and epigenetic control of early mouse development. *Curr Opin Genet Dev* 19:113–121
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K et al (2002) *Molecular biology of the cell*, 4th edn. Garland Science, New York
- Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF et al (2006) Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 8:532–538

- Bai L, Morozov AV (2010) Gene regulation by nucleosome positioning. *Trends Genet* 26:476–483
- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21:381–395
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE et al (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837
- Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. *Genes Dev* 23:781–783
- Bergsland M, Ramsköld D, Zaouter C, Klum S, Sandberg R et al (2011) Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* 25:2453–2464
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125:315–326
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A et al (2010) The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 28:1045–1048
- Bird A (2007) Perceptions of epigenetics. *Nature* 447:396–398
- Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T et al (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299:2877–2883
- Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F et al (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336:934–937
- Borgel J, Guibert S, Li Y, Chiba H, Schübeler D et al (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42:1093–1100
- Brenner C, Fuks F (2007) A methylation rendezvous: reader meets writers. *Dev Cell* 12:843–844
- Buecker C, Wysocka J (2012) Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet* 28:276–284
- Burgold T, Spreafico F, De Santa F, Totaro MG, Prosperini E et al (2008) The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. *PLoS One* 3:e3034
- Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S et al (2009) Unlocking the secrets of the genome. *Nature* 459:927–930
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW et al (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107:21931–21936
- Cui K, Zang C, Roh TY, Schones DE, Childs RW et al (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4:80–93
- Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C et al (2009) H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nat Struct Mol Biol* 16:777–781
- Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S et al (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9:166–175
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22:1775–1789
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T et al (2012) Landscape of transcription in human cells. *Nature* 489:101–108
- Dong X, Greven MC, Kundaje A, Djebali S, Brown JB et al (2012) Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol* 13:R53
- Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW et al (2006) Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol* 13:713–719
- Dovey OM, Foster CT, Cowley SM (2010) Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proc Natl Acad Sci U S A* 107:8242–8247
- ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1 % of the human genome by the ENCODE pilot project. *Nature* 447:799–816

- ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74
- Epsztejn-Litman S, Feldman N, Abu-Remaileh M, Shufaro Y, Gerson A et al (2008) De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat Struct Mol Biol* 15:1176–1183
- Erhardt S, Su IH, Schneider R, Barton S, Bannister AJ et al (2003) Consequences of the depletion of zygotic and embryonic enhancer of zeste 2 during preimplantation mouse development. *Development* 130:4235–4248
- Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49
- Farthing CR, Ficiz G, Ng RK, Chan CF, Andrews S et al (2008) Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet* 4:e1000116
- Fazio TG, Huff JT, Panning B (2008) An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134:162–174
- Feldman N, Gerson A, Fang J, Li E, Zhang Y et al (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8:188–194
- Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447:433–440
- Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F et al (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473:398–402
- Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A et al (2008) Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2:160–169
- Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ et al (2009) Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 460:863–868
- Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M (2011) Open chromatin in pluripotency and reprogramming. *Nat Rev Mol Cell Biol* 12:36–47
- Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG et al (2010) A map of open chromatin in human pancreatic islets. *Nat Genet* 42:255–259
- Grewal SI (2010) RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* 20:134–141
- Gu TP, Guo F, Yang H, Wu HP, Xu GF et al (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477:606–610
- Guttman M, Amit I, Garber M, French C, Lin MF et al (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK et al (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300
- Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK et al (2011) Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res* 21:1393–1409
- He YF, Li BZ, Li Z, Liu P, Wang Y et al (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333:1303–1307
- Hemberger M, Dean W, Reik W (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat Rev Mol Cell Biol* 10:526–537
- Ho L, Miller EL, Ronan JL, Ho WQ, Jothi R et al (2011) esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol* 13:903–913
- Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C et al (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* 104:829–838
- Huang HS, Matevosian A, Whittle C, Kim SY, Schumacher A et al (2007) Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. *J Neurosci* 27:11254–11262

- Issaeva I, Zonis Y, Rozovskaia T, Orlovsky K, Croce CM et al (2007) Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* 27:1889–1903
- Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334:194
- Ito S, Shen L, Dai Q, Wu SC, Collins LB et al (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333:1300–1303
- Jeffries MA, Dozmorov M, Tang Y, Merrill JT, Wren JD et al (2011) Genome-wide DNA methylation patterns in CD4⁺ T cells from patients with systemic lupus erythematosus. *Epigenetics* 6:593–601
- Jiang H, Shukla A, Wang X, Chen WY, Bernstein BE et al (2011) Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* 144:513–525
- Jones PA, Archer TK, Baylin SB, Beck S, Berger S et al (2008) Moving AHEAD with an international human epigenome project. *Nature* 454:711–715
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA et al (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467:430–435
- Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA et al (2006) The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* 8:285–292
- Kaneda R, Takada S, Yamashita Y, Choi YL, Nonaka-Sarukawa M et al (2009) Genome-wide histone methylation profile for heart failure. *Genes Cells* 14:69–77
- Kanhere A, Viiri K, Araújo CC, Rasaiyaah J, Bouwman RD et al (2010) Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol Cell* 38:675–688
- Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M et al (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 4:e1000242
- Kunath T, Arnaud D, Uy GD, Okamoto I, Chureau C et al (2005) Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development* 132:1649–1661
- Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D et al (2012) Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res* 22:1735–1747
- Kwon GS, Viotti M, Hadjantonakis AK (2008) The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev Cell* 15:509–520
- Lan F, Bayliss PE, Rinn JL, Whetstine JR, Wang JK et al (2007) A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449:689–694
- Lane N, Dean W, Erhardt S, Hajkova P, Surani A et al (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35:88–93
- Lee MG, Villa R, Trojer P, Norman J, Yan KP et al (2007) Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318:447–450
- Levasseur DN, Wang J, Dorschner MO, Stamatoyannopoulos JA, Orkin SH (2008) Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev* 22:575–580
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915–926
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M et al (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148:84–98
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F et al (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43:1091–1097
- Lim CY, Tam WL, Zhang J, Ang HS, Jia H et al (2008) Sall4 regulates distinct transcription circuitries in different blastocyst-derived stem cell lineages. *Cell Stem Cell* 3:543–554

- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W et al (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38:431–440
- Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K et al (2011) Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25:594–607
- Marks H, Kalkan T, Menafra R, Denissov S, Jones K et al (2012) The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149:590–604
- Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 7:29–59
- Matsui Y, Zsebo K, Hogan BL (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* 70:841–847
- Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H et al (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464:927–931
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–1195
- Mazzarella L, Jørgensen HF, Soza-Ried J, Terry AV, Pearson S et al (2011) Embryonic stem cell-derived hemangioblasts remain epigenetically plastic and require PRC1 to prevent neural gene expression. *Blood* 117:83–87
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J et al (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766–770
- Miao F, Smith DD, Zhang L, Min A, Feng W et al (2008) Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes. *Diabetes* 57:3189–3198
- Miao F, Chen Z, Zhang L, Liu Z, Wu X et al (2012) Profiles of epigenetic histone post-translational modifications at type 1 diabetes susceptible genes. *J Biol Chem* 287:16335–16345
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560
- Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S et al (2008) Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet* 82:696–711
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M et al (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* 113:631–642
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J et al (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30:755–766
- Morán I, Akerman I, van de Bunt M, Xie R, Benazra M et al (2012) Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in Type 2 diabetes. *Cell Metab* 16:435–448
- Mouse ENCODE Consortium (2012) An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol* 13:418
- Nashun B, Yukawa M, Liu H, Akiyama T, Aoki F (2010) Changes in the nuclear deposition of histone H2A variants during pre-implantation development in mice. *Development* 137:3785–3794
- Nazor KL, Altun G, Lynch C, Tran H, Harness JV et al (2012) Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* 10:620–634
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E et al (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489:83–90
- Niwa H, Toyooka Y, Shimamoto D, Strumpf D, Takahashi K et al (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* 123:917–929
- Ng RK, Dean W, Dawson C, Lucifero D, Madeja Z et al (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol* 10:1280–1290

- Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257
- O'Neill LP, VerMilyea MD, Turner BM (2006) Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat Genet* 38:835–841
- Pan G, Tian S, Nie J, Yang C, Ruotti V et al (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* 1:299–312
- Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The Polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27:3769–3779
- Pasini D, Hansen KH, Christensen J, Agger K, Cloos PA et al (2008) Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* 22:1345–1355
- Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP et al (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464:306–310
- Pei B, Sisu C, Frankish A, Howald C, Habegger L et al (2012) The GENCODE pseudogene resource. *Genome Biol* 13:R51
- Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y et al (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139:1290–1302
- Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* 28:1057–1068
- Probst AV, Santos F, Reik W, Almouzni G, Dean W (2007) Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. *Chromosoma* 116:403–415
- Puschendorf M, Terranova R, Boutsma E, Mao X, Isono K et al (2008) PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* 40:411–420
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA et al (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470:279–283
- Raiber EA, Beraldi D, Ficiz G, Burgess H, Branco MR et al (2012) Genome-wide distribution of 5-formylcytosine in ES cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* 13:R69
- Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S et al (2011) Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* 7:e1002300
- Ramsahoye BH, Biniszkievicz D, Lyko F, Clark V, Bird AP et al (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* 97:5237–5242
- Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293:1089–1093
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S et al (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463:237–240
- Sakaue M, Ohta H, Kumaki Y, Oda M, Sakaide Y et al (2010) DNA methylation is dispensable for the growth and survival of the extraembryonic lineages. *Curr Biol* 20:1452–1457
- Sandovici I, Leppert M, Hawk PR, Suarez A, Linares Y et al (2003) Familial aggregation of abnormal methylation of parental alleles at the IGF2/H19 and IGF2R differentially methylated regions. *Hum Mol Genet* 12:1569–1578
- Sandovici I, Smith NH, Ozanne S, Constância M (2008) The dynamic epigenome: the impact of the environment on epigenetic regulation of gene expression and developmental programming. In: Tost J (ed) *Epigenetics*. Caister Academic, Norfolk, pp 343–369
- Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S et al (2011) Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets. *Proc Natl Acad Sci U S A* 108:5449–5454

- Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ et al (2010) Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* 12:853–862
- Santos F, Hendrich B, Reik W, Dean W (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* 241:172–182
- Santos F, Peters AH, Otte AP, Reik W, Dean W (2005) Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol* 280:225–236
- Santos J, Pereira CF, Di-Gregorio A, Spruce T, Alder O et al (2010) Differences in the epigenetic and reprogramming properties of pluripotent and extra-embryonic stem cells implicate chromatin remodelling as an important early event in the developing mouse embryo. *Epigenetics Chromatin* 3:1
- Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE et al (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419:407–411
- Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. *Nature* 489:109–113
- Sarmiento OF, Digilio LC, Wang Y, Perlin J, Herr JC et al (2004) Dynamic alterations of specific histone modifications during early murine development. *J Cell Sci* 117:4449–4459
- Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. *Genome Res* 22:1748–1759
- Schmitz SU, Albert M, Malatesta M, Morey L, Johansen JV et al (2011) Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J* 30:4586–4600
- Seenundun S, Rampalli S, Liu QC, Aziz A, Palii C et al (2010) UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. *EMBO J* 29:1401–1411
- Senner CE, Krueger F, Oxley D, Andrews S, Hemberger M (2012) DNA methylation profiles define stem cell identity and reveal a tight embryonic-extraembryonic lineage boundary. *Stem Cells* 30:2732–2745
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A et al (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484:339–344
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F et al (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480:490–495
- Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M et al (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* 9:1428–1435
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465–476
- Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T et al (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev* 16:1779–1791
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935
- Talbert PB, Henikoff S (2010) Histone variants—ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* 11:264–275
- Tan M, Luo H, Lee S, Jin F, Yang JS et al (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146:1016–1028
- Tanaka S, Kunath T, Hadjantonakis AK, Nagy A, Rossant J (1998) Promotion of trophoblast stem cell proliferation by FGF4. *Science* 282:2072–2075
- Teo AK, Arnold SJ, Trotter MW, Brown S, Ang LT et al (2011) Pluripotency factors regulate definitive endoderm specification through *omesodermin*. *Genes Dev* 25:238–250
- The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT et al (2012) The accessible chromatin landscape of the human genome. *Nature* 489:75–82

- Torres-Padilla ME, Bannister AJ, Hurd PJ, Kouzarides T, Zernicka-Goetz M (2006) Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol* 50:455–461
- Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M et al (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11:805–814
- van der Heijden GW, Dieker JW, Derijck AA, Muller S, Berden JH et al (2005) Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. *Mech Dev* 122:1008–1022
- Volkmar M, Dedeurwaerder S, Cunha DA, Ndlovu MN, Defrance M et al (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J* 31:1405–1426
- Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K et al (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 6:153–166
- Wang H, Maurano MT, Qu H, Varley KE, Gertz J et al (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* 22:1680–1688
- Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* 41:246–250
- Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE (2010) A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140:99–110
- Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V et al (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2:241
- Wu H, D'Alessio AC, Ito S, Xia K, Wang Z et al (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473:389–393
- Wu Q, Chen X, Zhang J, Loh YH, Low TY et al (2006) Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem* 281:24090–24094
- Yagi R, Kohn MJ, Karavanova I, Kaneko KJ, Vullhorst D et al (2007) Transcription factor TEAD4 specifies the trophoctoderm lineage at the beginning of mammalian development. *Development* 134:3827–3836
- Yildirim O, Li R, Hung JH, Chen PB, Dong X et al (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147:1498–1510
- Zaret KS, Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 25:2227–2241
- Zentner GE, Tesar PJ, Scacheri PC (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* 21:1273–1283
- Zhao XD, Han X, Chew JL, Liu J, Chiu KP et al (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* 1:286–298
- Ziller MJ, Müller F, Liao J, Zhang Y, Gu H et al (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* 7:e1002389

Chapter 3

X-Chromosome Inactivation

Wendy P. Robinson, Allison M. Cotton, Maria S. Peñaherrera,
Samantha B. Peeters, and Carolyn J. Brown

Abstract The dimorphism of the sex chromosomes has led to some of the most dramatic epigenetic phenomena known in order to achieve dosage compensation between the sexes. Mammalian X-chromosome inactivation (XCI) requires the differential treatment of two essentially identical chromosomes in the same nuclear environment. XCI has thus been the subject of considerable study in mouse as a paradigm for epigenetic choice, although less is known about the timing and initial events of XCI in humans and other species. XCI, as can be visualized by the spots on a calico cat, is also a dramatic example of the stability of epigenetic silencing, since the inactivation state of an X chromosome (X) is faithfully inherited through subsequent somatic cell divisions. Studies to understand the layering of epigenetic modifications that result in such stable silencing have been reviewed elsewhere, and in this review we focus instead on the translation of our growing understanding of the epigenetic phenomena of XCI to human disease.

X-linked disease is epitomized by an excess of affected males, but the characterization as dominant or recessive belies the complexity of the contribution of XCI. Notably, whether or not X-linked disease is apparent in females is considerably impacted by the extent of skewing of XCI in the individual. Furthermore, the

W.P. Robinson • M.S. Peñaherrera
Department of Medical Genetics, Molecular Epigenetics Group,
University of British Columbia, Vancouver, BC, Canada V6T 1Z3

Child & Family Research Institute, 950 West 28th Avenue,
Vancouver, BC, Canada V5Z 4H4
e-mail: wrobinson@cfri.ca; maria.penaherrera@ubc.ca

A.M. Cotton • S.B. Peeters • C.J. Brown (✉)
Department of Medical Genetics, Molecular Epigenetics Group,
University of British Columbia, Vancouver, BC, Canada V6T 1Z3
e-mail: allison.cotton@ubc.ca; s.peeters@alumni.ubc.ca; carolyn.brown@ubc.ca

unique biology of the sex chromosomes impacts the likelihood of X-linked disease in females due to de novo mutation rates. In addition, XCI does not result in complete dosage equivalence between males and females; however, the extent to which there are sex, and even interindividual differences due to XCI has not yet been well elucidated. Overall, while it is the Y chromosome (Y) that determines sex, the X contributes in a complex fashion to the sex differences in disease frequency and severity.

3.1 Mosaicism and Skewing in XCI

A consequence of X-chromosome inactivation (XCI) is that females are mosaic for the expression of the X-linked genes subject to inactivation. This confounds the classic concept of “dominant” or “recessive” inheritance, as the ability of a normal allele to compensate for a mutation-carrying allele will depend on whether the gene product functions cell autonomously, as well as on the interactions between normal and mutation-expressing cells and their products. Furthermore, despite XCI being a random process, most tissue samples will show some detectable degree of bias (skewing) towards inactivation of either the maternal or paternal X. The proportion of cells expressing the mutant copy can vary from 0 to 100 %, and this proportion can vary between tissues sampled and between sites within a tissue. Skewed expression of one or the other gene copy can occur due to: biased initial inactivation; restricted precursor cell pool size at the time of XCI or during early development; and/or active selection (see Fig. 3.1). It is important to understand the dynamics of random and nonrandom XCI during development to be able to interpret the potential phenotype of an X-linked mutation in a female carrier. Assessment of XCI skewing in female carriers of an X-linked mutation can be useful in the clinical setting in certain situations, but needs to be interpreted with considerable caution.

3.1.1 Causes of Skewing of XCI

3.1.1.1 Skewing Related to Choice or Biased Initial Inactivation

A primary bias in the choice of the X to remain active could arise due to mutations in the X-inactivation center, including the X-inactive specific transcript (*XIST*) involved in the initiation of XCI (see Minks et al. 2008). While primary skewing of XCI is observed in mouse (Cattanach et al. 1969), in humans it has been difficult to discriminate this mechanism from early selection against cells inactivating one or the other X due to mutation. Rare mutations in the *XIST* minimal promoter (Plenge et al. 1997; Tomkins et al. 2002) have been associated with skewed XCI or failure to inactivate the mutated X, and also have been reported to impact the binding ability of the boundary factor CCCTC-binding factor (zinc finger protein) CTCF

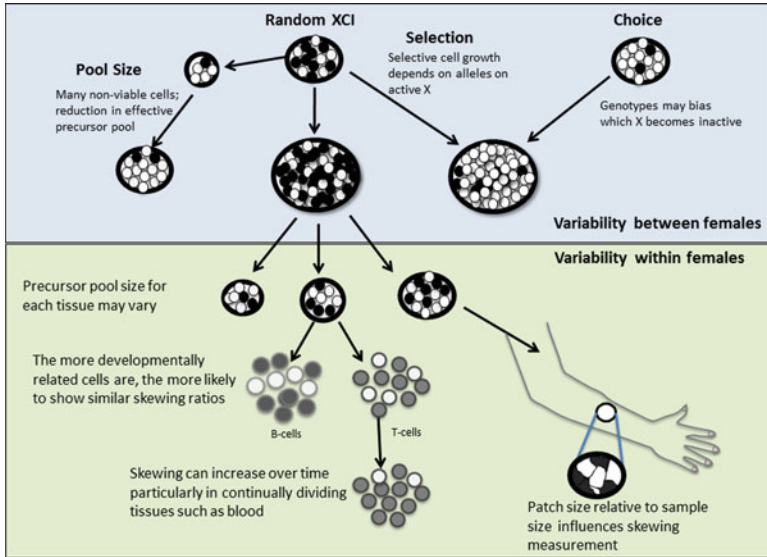


Fig. 3.1 The origins of skewed XCI and sources of variability within and between females. Females may show nonrandom XCI due to a primary bias in choice of X to inactivate; due to a secondary growth advantage of one population; or due to a reduction in the size of the pool of cells contributing to the embryo. Within any female there can be further variability in XCI skewing between tissues, with age, or due to sampling

(Pugacheva et al. 2005), although genome-wide surveys have not shown CTCF enrichment at this location in somatic cells. Such mutations were not observed in any of 66 women presenting with skewed XCI of unknown etiology (Pereira and Zata 1999), so are not a common cause of skewing. A pedigree with unexpected hemophilia A expression in females showed linkage to a region within Xq25 that overlapped similar regions identified in previous studies (Cau et al. 2006; Naumova et al. 1998). Within this region, stromal antigen 2 (*STAG2*) was identified as a candidate gene for a role in XCI choice in humans due to its function as part of the cohesin complex and interaction with CTCF (Renault et al. 2011). Overall, it remains to be determined whether there are variants affecting the initial choice of X to inactivate in humans.

3.1.1.2 Skewing Due to Selection

Selection as a mechanism for nonrandom XCI is based on differential cellular growth or cell survival following random XCI. The clearest examples are provided by structural chromosomal abnormalities in which preferential survival of those cells with the abnormal X inactivated is observed. In balanced X;autosome translocations ($t(X;A)$) selection typically favors those cells where the normal X is inactive. In contrast, unbalanced $t(X;A)$ typically show preferential

inactivation of the t(X;A) (reviewed in Leppig and Disteché 2001). The degree of skewing depends on the timing and efficiency of selection; however, the end result generally favors the balanced expression of autosomal and X-linked genes (Leppig and Disteché 2001; Schluth et al. 2007).

Although selection is an attractive mechanism to explain skewed XCI, most women found to have extremely skewed XCI have normal chromosomes. Among 45 women with skewed XCI (>80 %) there was not an increase in copy number variants when screened by high resolution array for cryptic X deletions and duplications (Jobanputra et al. 2012). Only a single 5.5-Mb deletion was identified in one female with skewed XCI, suggesting that cryptic chromosomal abnormalities may only rarely account for skewing.

Mutations within X-linked genes can also contribute to skewed XCI. While random XCI is generally observed in female carriers for most X-linked gene disorders, there are conditions where the mutation results in skewed XCI with preferential expression of the normal allele of the gene or, more rarely the mutant allele, as is seen for adrenoleukodystrophy (reviewed in Orstavik 2009; Salsano et al. 2012). Depending on the mutation, skewing can occur shortly after XCI and thus be constitutional, or may be limited to specific tissues in which the gene is expressed. For example, mutations in forkhead box P3 (*FOXP3*) cause dysfunctional T-cells leading to immune dysregulation polyendocrinopathy enteropathy (IPEX syndrome) in males, yet carrier women are usually unaffected. Skewed XCI in carrier females is limited specifically to the CD4⁺CD25^{hi} regulatory T-cells, despite random XCI being observed in naive and memory CD4⁺ T-cells (Di Nunzio et al. 2009). Therefore, evaluating skewing of XCI in a sample of peripheral blood lymphocytes from such cases will not be informative for identifying the presence of skewing in the clinically relevant subset of cells that are responsible for the clinical manifestations of disease.

3.1.1.3 Skewing Related to Precursor Pool Size

In the absence of selection, the distribution of skewing in a population of tissue samples depends on (1) the number of cells present in the precursor pool at the time an X is initially marked to be inactivated, (2) the number of cells from this precursor pool which contribute to the development of that particular tissue, and (3) the number of cell descendants which remain closely associated (patch size) in the tissue relative to the sample size analyzed. The skewing values in a population of samples from blood in newborns are largely normally distributed and consistent with derivation from a precursor pool of less than 16 cells (Amos-Landgraf et al. 2006). Similarly, the number of precursors contributing to blood was estimated at 4–5 precursors, whereas the less extreme XCI values in buccal samples suggested a precursor pool size of 16 cells (Monteiro et al. 1998). These numbers assume biases result only from stochastic forces and may be underestimates if selection plays a significant role.

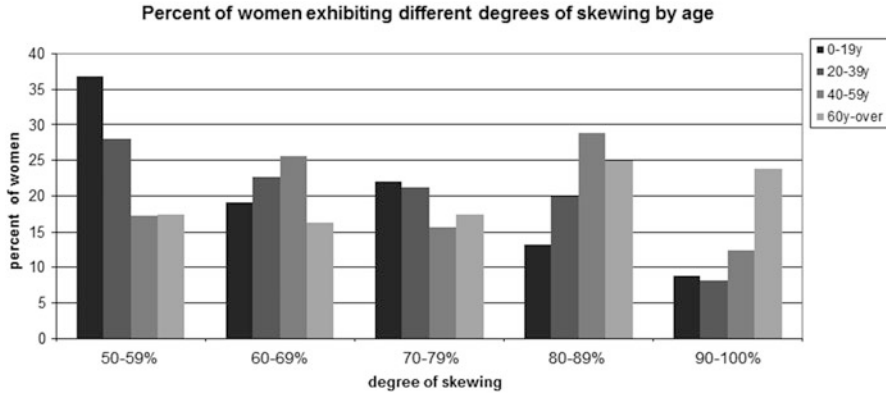


Fig. 3.2 The effect of age on extent of skewed XCI in blood. Older females show more nonrandom patterns of XCI. $N = 68$ for 0–19 years; $N = 146$ for 20–39 years; $N = 121$ for 40–59 years; $N = 80$ for 60–89 years

Precursor pool size may vary between individuals, and in some situations the embryo may derive from a highly reduced number of these precursors. For example, skewed XCI is frequently seen in newborns from pregnancies complicated by confined placental mosaicism, a situation where significant levels of trisomy are found exclusively in the placenta. Skewed XCI is presumably the result of trisomic cells being present in the embryonic precursor pool at the time of XCI that do not contribute significantly to the differentiated embryo (Lau et al. 1997). Similarly, completely skewed XCI is increased in Prader-Willi syndrome patients with uniparental disomy, a situation thought to frequently arise in association with confined placental trisomy (Lau et al. 1997; Butler et al. 2007).

3.1.1.4 Skewing Increases with Age

Numerous studies have shown that skewing gradually increases with age in samples of peripheral blood (Busque et al. 1996; Sharp et al. 2000; Hatakeyama et al. 2004; Knudsen et al. 2007; Bolduc et al. 2008). Based on our own data for 415 healthy females, the proportion of extreme skewing ($>90\%$) is less than 10% at younger ages but nearly 25% in women over 60 years of age (Fig. 3.2). The distribution of skewing varies from study to study and may be subject to lab and assay-based differences (methodologies are discussed below). Over a 2-year period the level of skewing in peripheral blood samples is remarkably stable at all ages (van Dijk et al. 2002) likely reflecting the stability of hematopoietic stem cell populations over limited time frames. It is probable that the increase in skewing over time is a combination of a slight proliferative advantage of cells that have inactivated one X over the other as well as random stochastic loss of stem cell precursors (Christensen et al. 2000). Interestingly, neutrophils showed a greater degree of skewing ($>90\%$) in elderly women than did T-cells (33% vs. 9%) (Gale and Linch 1998).

Blood is typically used to assess skewing as it is more readily accessible than other tissue sources; however, caution is needed as the skewing in the tissue of interest to a specific disease process may differ from blood (see Fig. 3.1). Skewing in different somatic tissues is correlated (Sharp et al. 2000; Bolduc et al. 2008; Bittel et al. 2008), but the proportion of samples showing moderate or extreme skewing tends to be less in skin and muscle (Gale et al. 1994) or buccal cells (Bolduc et al. 2008). Furthermore, this correlation decreases with age, suggesting either differences in selection or stochastic processes affecting skewing ratios over time in different tissues.

3.1.1.5 Skewing in Twins

There is little evidence that monozygotic (MZ) twin pairs show skewed XCI more often than dizygotic twins (Monteiro et al. 1998). There is, however, a significant correlation in degree and direction of XCI skewing between MZ twin pairs dependent on the developmental timing of twinning. Monoamniotic, monochorionic twin pairs display the highest correlation in skewing (Monteiro et al. 1998; Chitnis et al. 1999) reflecting their relatively later derivation from a common pool of cells after the commitment to XCI. Nonetheless, there can be substantial differences between XCI skewing in monozygotic twins, particularly dichorionic MZ pairs, and this may account for some phenotypic discordance in otherwise genetically identical pairs. MZ twin discordance in the presentation of X-linked disorders such as Fragile X syndrome and Duchenne muscular dystrophy (DMD) (Richards et al. 1990) has been attributed to differing XCI ratios.

3.1.1.6 Is Skewing Heritable?

Since skewing can reflect all the processes discussed above (and outlined in Fig. 3.1); it is challenging to determine whether skewed XCI is heritable in humans. Clearly transmission of X rearrangements or mutations that are accompanied by skewed XCI will lead to familial skewed XCI, but evidence for heritability in the general population is limited. In a study of 38 control families, there was a correlation in skewed XCI between sister–sister pairs but not in mother–daughter pairs (Naumova et al. 1998). A study of over 500 mother–neonate pairs also showed no correlation between degree of skewing or incidence of extreme skewing (Bolduc et al. 2008), suggesting that genetics has little influence on degree and direction of skewing in the general population.

A separate question is whether increased skewing of XCI with age is a heritable trait. Such a tendency could explain the correlation in skewing between adult siblings but not when newborns are compared to older relatives. The concordance in skewing between elderly twin pairs was used to infer that acquired skewing with age results from a small selective advantage of cells with one or the other X active, while stochastic events and depletion of the

stem cell pool may contribute to acquired skewing to some smaller degree (Kristiansen et al. 2005; Vickers et al. 2001). Interestingly, the offspring (aged ~60–80) of centenarians exhibited less XCI skewing than comparably aged females of non-long-lived parents (Gentilini et al. 2012), possibly due to reduced turnover of the hematopoietic stem cell pool, as they also have a lower incidence of disease. While intriguing, the study sample was relatively small and the cause of this association remains to be shown. Overall, it appears that the initial skewing ratios observed in newborns are largely determined by stochastic effects, but that over time subtle genetic differences can affect both the degree and direction of skewing, at least in blood.

3.1.2 Skewing of XCI: Impact on Disease

Most sex-linked disorders cannot readily be classified as dominant or recessive (see also Dobyns et al. 2004), since males are hemizygous, and female heterozygotes are generally mosaics of two different cell populations due to XCI. While it is difficult to classify X-linked disorders into clear patterns of inheritance it is perhaps useful to consider the following (see Fig. 3.3): (1) recessive-like, in which females are generally not affected; (2) dominant-like, in which females generally are affected; and (3) disease caused by mosaic XCI in females with males generally unaffected.

3.1.2.1 Recessive-Like: Males Affected and Generally Unaffected Females

Heterozygous females may not manifest an X-linked disorder when they are near random in their XCI pattern because the mosaic expression of the normal allele is sufficient to prevent disease. An example is the metabolic cooperation between cells with the active normal or mutant allele for mucopolysaccharidosis type II (Hunter syndrome), in which the deficiency is corrected via cell to cell transfer (Migeon 2006, and see Orstavik 2009 for additional examples). In other disorders females are unaffected because of a tendency towards a skewed pattern of XCI favoring the normal allele being kept active. Such selection may be evident in several different tissues as seen for alpha-thalassemia X-linked intellectual disability syndrome (ATRX) (Gibbons et al. 1992), or may be tissue specific as in the IPEX syndrome discussed above.

In this category of X-linked disorders, females may sometimes show limited signs or be affected with the disease for a variety of reasons. First, 45,X females (Turner syndrome) can be affected as the dosage of X-linked genes is equivalent to that of hemizygous affected males, as has been observed in an affected 45,X girl with DMD (Chelly et al. 1986). Second, unfortunate lyonization, or extreme skewing can result in the mutant allele being preferentially active. Balanced

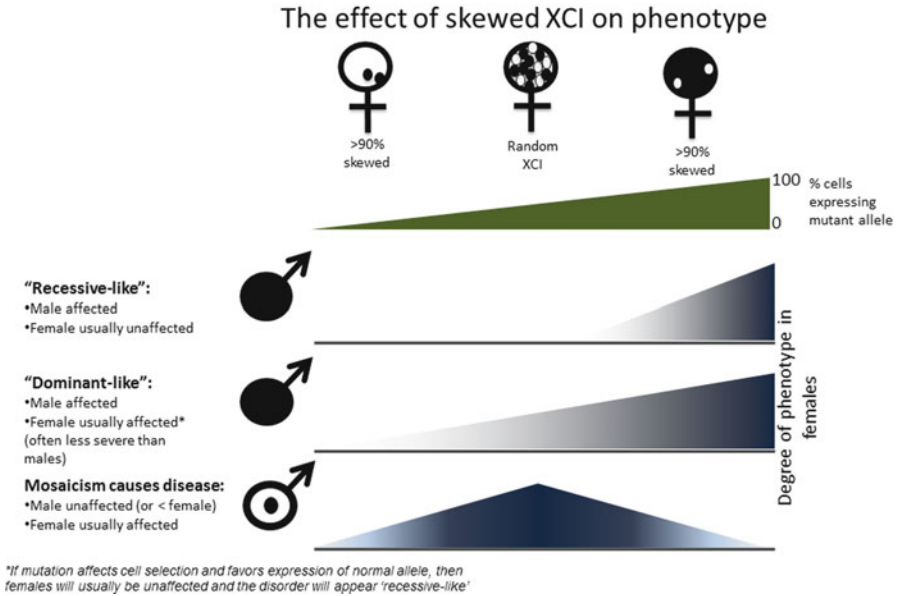


Fig. 3.3 The effect of skewed XCI on phenotype. For cell-autonomous traits, females will be mosaics; however, the extent of phenotypic expression will depend on the skewing of XCI in the female. For most X-linked disorders the hemizygous male will manifest the trait, while females will show the trait only when heavily skewed (recessive-like) or unless favorably skewed (dominant-like). In rare cases the presence of mosaicism causes the disease in which case males will be spared

t(X;A)s can disrupt a gene at the site of breakpoint and also result in skewed XCI, leading to manifestation of disorders such as DMD (Boyd and Buckle 1986). Interestingly, it has been suggested that increased nonrandom XCI with age may lead to loss of heterozygosity and potentially drive the onset of some X-linked disorders in elderly carrier women. This was suggested as a possible mechanism for glucose-6-phosphate dehydrogenase (*G6PD*) deficiency (Au et al. 2006) or late onset X-linked sideroblastic anemia (Cazzola et al. 2000).

3.1.2.2 Dominant-Like: Males Affected or Male Lethal; and Females Often Affected

Dominant-like disorders are those in which females and males generally show signs of disease. Typically males are more severely affected than females and in numerous cases the disorder is incompatible with survival in hemizygous males such that only affected females are observed. In some such disorders, such as focal dermal hypoplasia (Goltz-Gorlin syndrome) (Clements et al. 2009), there is surprisingly no clear correlation between XCI and phenotype. Occasionally, skewing towards inactivation of the mutant allele, or fortunate lyonization, will result in unaffected

female carriers (e.g., Rett syndrome; Dayer et al. 2007), while predominant inactivation of the normal allele will result in more severely affected females (e.g., Fabry disease Dobrovolny et al. 2005). Interestingly, rare exceptions to the male lethal phenotype of some conditions have been reported. These are often individuals with supernumerary Xs or having somatic mosaicism for the mutation; reports of this include males with Rett syndrome (Smeets et al. 2012), Aicardi syndrome (Hopkins et al. 1979), and focal dermal hypoplasia (Wang et al. 2007).

3.1.2.3 Mosaicism Dependent Phenotype: Females Affected More Severely than Males

Paradoxically, for some X-linked diseases heterozygous females have greater disease severity than hemizygous males. For craniofrontonasal syndrome, which is caused by mutations of ephrin-B1 (*EFNB1*) a marker of tissue boundary formation, it has been hypothesized that the patchwork loss of the gene expression in females may disturb tissue boundary formation (Twigg et al. 2004). In hemizygous males an alternative mechanism, perhaps due to the promiscuity of the ephrin ligand/receptor system, maintains the proper boundary formation; but in the female carriers mosaicism for the mutation causes disordered growth. The coexistence of normal and mutant allele products is suggested to lead to “cellular interference” and abnormal phenotype (Wieacker and Wieland 2005). A similar mechanism has been suggested for early infantile epileptic encephalopathy (Ryan et al. 1997). It is worth considering whether the presence of mosaicism for X-linked polymorphisms could contribute to other disorders with female preponderance. For example, if there are patches of the brain that are genetically distinct by virtue of expressing one or the other X this could affect the manner in which those patches of cells communicate with one another.

3.1.2.4 Skewed XCI and Complex Disorders for Which the Cause Is Unknown

Increased skewing is observed in some autoimmune disorders affecting predominantly women, including scleroderma (Ozbalkan et al. 2005), autoimmune thyroid disease (Ozelik et al. 2006), and juvenile arthritis (Uz et al. 2009), but not in others such as systemic lupus erythematosus. Intriguingly, monosomy X, which may often be mosaic, is also associated with autoimmunity, leading to the suggestion that a breakdown of self-tolerance may result from the loss of mosaic XCI in T-cells over time (reviewed in Ozelik et al. 2006). As the X has a large number of genes and microRNAs affecting immune function, mutations in a number of which lead to autoimmune disease (Bianchi et al. 2012), loss of heterozygosity for X-linked mutations or polymorphisms that affect immune function could also be playing a role in some autoimmune disorders.

Patients exhibiting skewed XCI and autoimmune thyroiditis showed an increased rate of pregnancy loss (Ozcelik et al. 2006). Furthermore, the X is over-represented for genes involved in reproductive function. Skewed XCI has been associated with recurrent miscarriage (RM, generally defined as three or more consecutive pregnancy losses) in several studies (Lanasa et al. 1999; Uehara et al. 2001; Beever et al. 2003b). A similar association was not found with the less strict criteria of two miscarriages, not consecutive, used in two other studies (Hogge et al. 2007; Warburton et al. 2009). In one large pedigree ascertained for increased miscarriage a deletion on the X was identified (Pegoraro et al. 1998). If mutations affecting male survival were a common explanation for skewed XCI in women experiencing RM a preponderance of karyotypically normal males among the RM would be anticipated, which has not been observed (Stephenson et al. 2002). A meta-analysis supported an association between skewed XCI and RM (Su et al. 2011), though this association remains controversial. Skewed XCI has also been associated with primary premature ovarian failure (POF) (Bretherick et al. 2007; Sato et al. 2004). Skewing in these females might reflect restricted embryo precursor size leading to compromised growth of germ cells, or the presence of X-linked mutations as POF is associated with a number of X rearrangements implicating several gene regions on the X (Toniolo and Rizzolio 2007).

3.1.3 Assessment of Skewing

Since skewing of XCI can impact the phenotype of female carriers, XCI assays are widely used diagnostic tools for these X-linked conditions. However, their reliable application in clinical medicine requires clear definitions of the phenotype, correction for possible age-related biases, corroboration of the phenotype-XCI skewing associations in independent datasets and specific tissues, awareness of the existing limitations of information available (particularly for rare conditions) and the use of reliable assays that are in linkage disequilibrium with the gene/region of interest.

Numerous assays used for determining XCI patterns have been described. These include electrophoretic assays using protein isoforms (G6PD); RFLPs (e.g., phosphoglycerate kinase 1 (*PGKI*)); or STR-based assays using PCR following methylation-sensitive enzyme digestion (e.g., androgen receptor (*AR*), fragile X mental retardation 1 (*FMR1*), monoamine oxidase A (*MAOA*), zinc finger MYM-type 3 (*ZNF261*), zinc finger DHHC-type containing 15 (*ZDHHC15*), SLIT and NTRK-like family member 4 (*SLITRK4*) and proprotein convertase subtilisin/kexin type 1 inhibitor (*PCSKIN*)). Each is challenged by technical issues such as incomplete digestion, microsatellite amplification stutter and biases that complicate the quantification of the products and can result in discordance between their results (reviewed in Beever et al. 2003a; Bertelsen et al. 2011).

The most widely used assay in clinical practice examines DNA methylation at the *AR* gene. This assay is informative in >90 % of females and evaluates the methylation pattern of two *HpaII* restriction sites adjacent to a polymorphic CAG

repeat in the 1st exon of *AR* in Xq12 (Allen et al. 1992). Restriction enzyme independent assays have been designed to interrogate the same region using methylation-specific PCR (M-PCR) following sodium bisulfite modification (Kubota et al. 1999). An expression-based *AR* assay that measures allelic expression has been shown to largely correlate with the methylation-based assay (Busque et al. 1994) and with other transcription-based assays (Bolduc et al. 2008). Additional assays based on the quantitative expression of polymorphic X-linked genes, also known as transcriptional clonality assays (qTCAs), have been developed but require access to RNA. While some studies have shown highly concordant results (Busque et al. 2009), there are reports of discordance between qTCAs and *AR* (methylation and expression) results that have been attributed to highly variable methylation of the *AR* gene in granulocytes (Swierczek et al. 2012). Discordance between assays is a concern that needs to be addressed, as it is yet not clear which assay provides the most accurate reflection of XCI patterns.

An important final consideration is that tissue specificity of XCI patterns (Gale et al. 1994) limits the conclusions that can be established based on the assessment of a single tissue, highlighting the need for appropriate tissue and age-matched controls in all XCI studies. In clinical practice, skewing is typically measured in peripheral blood for its accessibility; however, this tissue is not necessarily the most informative and relevant in certain conditions.

3.1.4 Skewed XCI to Assess Clonal Patch Size

XCI has long been used to assess clonality in tumors and other types of lesions; however, this analysis can be complicated by several considerations (Chen and Prchal 2007). The tumor may have incorporated cells of differing origins, such as vascular or inflammatory cells, and thus appear to be polyclonal. Alternately, there may be a subsequent expansion of a dominant clone even if the lesion was originally polyclonal in origin. Furthermore, due to the stability of XCI in cells descended from a common precursor, there is a natural patchiness to XCI patterns; thus an important consideration for clonality analysis in tumors is the patch size of the cells from which such lesions/tumors may arise. For example, a neoplastic origin of atherosclerotic lesions has been argued based on the observation of nonrandom XCI in the lesion. However, studies of XCI suggest that human arteries grow by expansion of smooth muscle cell clones with little mixing with adjacent clones, resulting in a patch length often >4 mm, and suggesting that plaques may simply arise from preexisting developmentally normal clones of cells (Chung et al. 1998).

Assays of XCI skewing estimate the patch size of epidermis to reflect an estimated 20–350 basal cells (Asplund et al. 2001). This implies a much finer scale substructure than reflected by the hyper- and hypo-pigmented skin patches seen to follow the lines of Blaschko in X-linked incontinentia pigmenti that had been proposed to represent XCI clonal patterns. Utilizing XCI to estimate clonal patch size, epidermis samples of 2 mm diameter were generally clonal, while

patches of 4–5 mm typically displayed a mixed XCI status (Chaturvedi et al. 2002). A monoclonal origin of individual crypts in both colon and small intestine was demonstrated by G6PD staining in multiple tissues from a heterozygote (Novelli et al. 2003). In colon, clusters of up to 450 adjacent crypts showed a similar staining pattern, while in small intestine a much smaller patch size was observed, with the epithelial lining of the villi sometimes showing mixed staining patterns. Both the myoepithelial and epithelial cells from individual breast ducts were of the same monoclonal origin and thus breast tumors, which tend to arise from individual ducts, would be expected to show skewed XCI even if having a multicellular origin (Novelli et al. 2003; Diallo et al. 2001). While thyroid follicles may be polyclonal origin (Novelli et al. 2003), monoclonal patch size in thyroid tissue was estimated at 48–128 mm² or $1\text{--}4 \times 10^5$ cells (Jovanovic et al. 2003).

Endometrial tissue shows dramatic regenerative capacity during the menstrual cycle with new proliferation deriving from stem cells within the endometrial glands. Glands located within an ~ 1 mm² area generally showed the same inactive X, whereas samples spanning >2 mm² showed a mixed pattern of XCI (Tanaka et al. 2003). Nonetheless, when paired 4–9 mm³ samples were analyzed from two distinct locations in endometrium, myometrium and cervix derived from tissues of 11 hysterectomies, a correlation in skewing measurements between the paired sites was observed (Mutter et al. 1996). The distribution of skewing was used to infer that each of these tissues derive from a stem cell pool of 10–12 cells and that one sample can be representative of the average skewing in the tissue as a whole.

Determining the relationship between skewed XCI in different tissues and in different samples from the same tissue provides a molecular assessment of the developmental history of a particular tissue that can then be used to understand how genetic and epigenetic variation arises in development. For example, the mean levels of XCI skewing for placental amnion and chorion are correlated, suggesting a common developmental origin from inner cell mass derivatives subsequent to XCI, while average skewing in placental trophoblast was uncorrelated with either amnion or chorion, consistent with its origin from the trophectoderm of the blastocyst (Penaherrera et al. 2012). Furthermore, the patterns of XCI skewing in trophoblast were consistent with a monoclonal origin of individual chorionic villus trees and large patch size. In contrast, different sites of amnion taken from the same placenta showed a high degree of correlation consistent with a high degree of intermixing of cells and little “patchiness.” Overall, skewing of XCI can provide important information on the developmental origins of cell populations, but care must be taken to ensure that the normal mosaic patterns of clonality in females are adequately considered.

3.2 De Novo Mutations in X-Linked Disease

For counseling of families with X-linked disease, it is important to identify non-manifesting carriers. Skewing of XCI has sometimes been used to identify carriers within a family (reviewed in Puck and Willard 1998); however, as discussed above

there can be multiple causes for skewed XCI. Furthermore, new mutations arise at a substantial frequency; in the case of an X-linked disorder severe enough to reduce the reproductive fitness of males, then for the disease frequency to remain constant in the population there must be sufficient new mutations to “replace” the alleles lost. While in theory these de novo mutations could be of maternal origin, the high frequency of carrier mothers observed in Lesch-Nyhan disease led to the suggestion that the de novo mutation rate was higher in males (Francke et al. 1976). Genome sequencing has now demonstrated that the male mutation rate is elevated throughout the genome (Kong et al. 2012). Higher paternal mutation rates have been seen for hemophilia A (Leuer et al. 2001) and B (Green et al. 1999), adrenoleukodystrophy (Wang et al. 2011), and X-linked hypophosphatemic rickets (Durmaz et al. 2013). The excess of paternal X-linked mutations is both gene and mutation type specific. For example, in DMD and Rett syndrome point mutations are predominantly paternal, while deletions or insertions can be maternal in origin (Grimm et al. 1994; Zhang et al. 2011).

The presence of germ line or somatic mosaicism has important counseling implications that depend on the gene and mutation. Somatic mosaicism has been observed in 8–12 % of mothers of children with X-linked hemophilia but not adrenoleukodystrophy (Wang et al. 2011). Somatic mutations in tumor suppressor genes can contribute to the development of cancer and generally require two hits to inactivate both alleles; however, as the X is functionally hemizygous, only a single hit is required. There are a limited number of X-linked genes known to be recurrently mutated in cancer, including APC membrane recruitment protein 1 (*AMER1*), *ATRX*, *FOXP3*, and PHD finger protein 6 (*PHF6*), the latter of which has been reported to have somatic mutations preferentially on the paternal X (Van Vlierberghe et al. 2011). In general, one would anticipate that XCI would result in both males and females having a single active allele subject to equivalent mutation probability. However, for those genes that escape XCI, females might show lower cancer frequency. With improved understanding of the genes that escape XCI such sexual dimorphisms could be better understood.

3.3 Gaps in Dosage Compensation

3.3.1 *Genes That Escape from XCI*

The unique evolutionary history of the sex chromosomes has driven the need for dosage compensation; but this compensation is incomplete, with approximately 15 % of human genes escaping from XCI (Carrel and Willard 2005). The eutherian X and Y diverged from each other approximately 150–160 mYa (reviewed in Livornois et al. 2012). Once the proto-Y obtained the sex-determining region Y (*SRY*) gene, step-wise decay of the Y ensued whenever recombination with the X was inhibited by genomic rearrangements such as inversions. This ratcheted loss of Y homology has resulted in “evolutionary strata” on the X (Lahn and Page 1999).

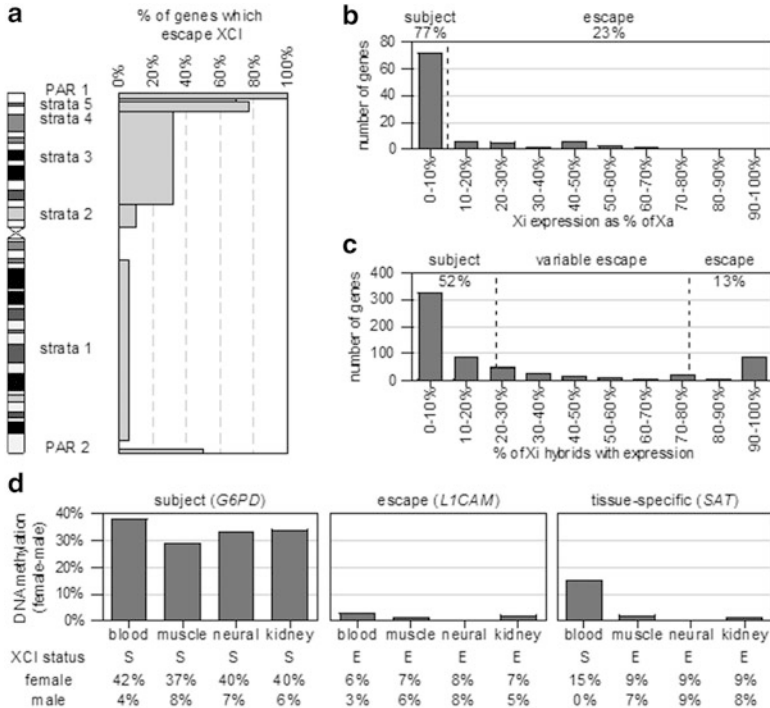


Fig. 3.4 Approximately 15 % of genes escape XCI. (a) The majority of genes identified to escape from XCI are located in regions of the X that are more recently diverged from the Y. (b) Using cSNPs to examine expression in females with nonrandom XCI, the level of expression from the inactive X was shown to be variable, but averages 42 % (Carrel and Willard 2005). (c) With somatic cell hybrids the number of escape genes was found to be lower, and for a substantial portion of the genes, expression was different between Xs in different hybrids. (d) DNA methylation differences between females and males at X-linked CpG-island promoters can be used to predict genes subject to XCI (e.g., *G6PD*), escaping from XCI (e.g., *L1CAM*) or escaping from XCI in only some tissues (e.g., *SAT*). S = subject to XCI, E = escapes XCI

The most recent stratum is the pseudoautosomal region (PAR), a region that still pairs and undergoes recombination between the X and Y during male meiosis. The boundary of the PAR is highly divergent between eutheria, and the human PAR1, which contains ~25 genes in 2.7 Mb of DNA, is smaller than most other eutheria except mouse (Ross et al. 2005).

Lyon first hypothesized XCI in 1961, and shortly thereafter suggested that genes in the PAR would not require dosage compensation (Lyon 1962). Indeed, all PAR1 genes examined to date escape XCI (Carrel and Willard 2005). Interestingly, two of the four PAR2 genes examined maintain dosage equivalence by silencing both the inactive X- and Y-linked genes (D'Esposito et al. 1996; De Bonis et al. 2006). As evolutionary time since divergence from the Y homolog increases, so does the probability that genes are subject to XCI (as shown in Fig. 3.4a). The escapees outside of the PARs include genes with functional Y homologues, such as zinc

finger protein, X-linked (*ZFX*) and ribosomal protein S4, X-linked (*RPS4X*), as well as genes with nonfunctional Y homologs, such as Kallmann syndrome 1 sequence (*KALI*), or without any apparent remaining Y homology such as carbonic anhydrase VB, mitochondrial (*CA5B*).

In mammals, the degradation of the Y not only led to dosage imbalance between males and females but also a potential halving of expression levels. While there has been some controversy, it appears that dosage compensation in eutheria involves both up-regulation of the active X in males and females and XCI to achieve dosage equivalency between the sexes. Comparisons of X and autosomal gene expression are complicated by the presence of considerably more tissue-specific genes on the X such as genes involved in reproduction; however, with the exclusion of such genes there is up-regulation of gene expression from the active X (see Deng et al. 2011 and references therein). A chromosomal control of up-regulation is suggested by the absence of up-regulation in all tissues, as dosage compensation is maintained in the gametes by not up-regulating the single X of sperm or two active Xs in oocytes (Nguyen and Disteche 2006). Dosage compensation is likely critical only for a subset of genes, in particular those whose products are involved in large protein complexes (Pessia et al. 2012).

3.3.2 Methodology for Determining XCI Status

It is becoming apparent that silencing of a gene by XCI is not an all or none epigenetic phenomena, but rather can be variable between individuals and tissues, as well as being nuanced in the level of expression from the inactive X. The sensitivity to detect such variabilities depends upon the approach used to determine inactivation status of a gene.

3.3.2.1 Expression Analysis in Heterozygous Females

As discussed, heterozygous carrier females can show mosaic expression, or selective skewing of XCI, both of which can be used as evidence that a gene is subject to XCI. Assessment of total RNA levels between males, females, and individuals with X aneuploidies have consistently identified some genes that escape XCI, but will also identify sex or aneuploidy-associated differences. To determine if expression is from one or both chromosomes, one can use RNA-FISH or, more commonly, expression of a heterozygous polymorphism in a female with skewed XCI. In such clonal populations, biallelic expression reflects expression from both the active and the inactive X. In the most extensive survey of biallelic expression to date, Carrel and Willard in 2005 evaluated 93 genes in clonal cell lines (see Fig. 3.4b) (Carrel and Willard 2005). While this is the most direct assay to detect expression from the inactive X, the need for an expressed SNP in a clonal cell population, or single-cell analysis, reduces the informativity of this approach.

Allelic expression imbalances will be quantitated by RNA sequencing if the cell population being examined is skewed for XCI, as has been demonstrated in mouse using cells from a cross between *Mus musculus* and *Mus spretus* to identify genes that escape XCI (Yang et al. 2010).

3.3.2.2 Expression Analysis in Somatic Cell Hybrids

The human inactive X can be isolated from the human active X in mouse/human somatic cell hybrids, allowing direct analysis of expression of human genes by human-specific RT-PCR (see Fig. 3.4c). Hybrids have been shown to lose localization of the XIST RNA that is essential for the initiation (Penny et al. 1996) but not maintenance (Brown and Willard 1994) of XCI; however, in the Carrel and Willard survey 91 genes were examined by both SNP expression and hybrid approaches and the inactivation status was comparable (Carrel and Willard 2005).

3.3.2.3 Assessment of Epigenetic Marks

Indirect assays for the XCI status of X-linked genes have been developed that rely upon the epigenetic marks that are associated with an active or inactive X. Analogous to the direct study of expression, the assessment of marks associated with gene activity needs to be combined with allelic differences in a clonal population of cells to identify whether there is monoallelic or biallelic presence of the mark. A number of X-linked promoters from genes escaping XCI were shown to have biallelic RNA polymerase II association (Kucera et al. 2011). In contrast, for marks that are associated with the silent allele, such as promoter DNA methylation, the presence of the mark can be considered as evidence for inactivation of the gene, without the need for polymorphisms and clonal cells or single-cell analysis. Thus, DNA methylation has been a very popular approach to examine XCI status of genes, with several groups having reported studies in a variety of tissues (Cotton et al. 2011; Sharp et al. 2011; Yasukochi et al. 2010). While expression or active mark studies are limited to genes expressed in the tissue examined, DNA methylation and other heterochromatic marks, such as H3K27me3 (Berletch et al. 2010), appear to be present even when the gene is not expressed in the assayed tissue. These marks provide the opportunity to study tissue-specific genes not assessed in the usual surveys that examine expression in fibroblasts or blood. DNA methylation, however, is generally only correlated with inactivation at genes with CpG-island promoters representing approximately 60–70 % of X-linked genes. Combining results from analyzing epigenetic marks with expression analyses will allow the most complete assessment of inactivation status for X-linked genes. However, the compilation of a catalog of inactivation status for X-linked genes is confounded by the variability that is now being revealed between different tissues and different individuals.

3.3.3 *Variability in XCI*

Rather than a dichotomous “on” or “off,” expression from the inactive X shows variability in the level of expression, the percent of individuals in which the gene is expressed, and the tissues in which the gene escapes inactivation. Thresholds have been set to establish XCI status in surveys; however, these variabilities need to be considered on a gene-by-gene basis in the clinic.

3.3.3.1 *Variability in Level of Expression from the Inactive X*

The relatively lower expression of a gene from the inactive X relative to the active X was a feature noted early in the study of genes escaping XCI (Migeon et al. 1982), and Carrel and Willard (Carrel and Willard 2005) established a “cutoff” of 10 % expression from the inactive X as the definition for “escape from XCI” (Carrel and Willard 2005) (see Fig. 3.4). As greater depth of quantitative data is acquired through approaches such as RNA-seq, the extent of this variability will be better quantified. It is likely that expression from the inactive X spans the spectrum from silent to active and it will be important to identify the biological point at which expression from the inactive X becomes relevant, because of increased expression in females, ability to avoid X-linked disease in females due to ongoing expression from the non-mutated X, or sensitivity or resilience of individuals with aneuploidies to disease. Depending on the approach used to identify expression from the inactive X, low level expression might reflect some cells expressing fully and others silencing completely. In the cases when RNA-FISH or single-cell PCR have been used to examine expression, it appears that cells have reduced expression from the inactive X (Carrel and Willard 1993); however, variability is also observed in which genes escape XCI in different tissues (see below).

3.3.3.2 *Variability Between Females in Escape from XCI*

A survey of expression in a panel of nine inactive X containing somatic cell hybrids showed that for many genes expression is observed in a subset of the hybrids. A cutoff of 2/9 (22 %) ((Carrel and Willard 2005); see Fig. 3.4c) was established for the designation of a gene as subject to XCI, seen for 52 % of genes, and 7/9 or more expressing hybrids was classified as “escape,” seen for 13 % of genes. Such variability, with a gene escaping inactivation in some, but not all, females has also been demonstrated using expressed SNPs (Anderson and Brown 1999; Carrel and Willard 1993; Kucera et al. 2011). Prediction of escape from XCI by DNA methylation also supports there being a substantial number of genes which escape XCI in some females (Cotton et al. 2011). This variable expression does not appear to be regulated at the level of the chromosome, as females who express one variable gene are not more likely to express another one. While it is unknown what local

features lead to variable escape of XCI, hypoacetylation of the metallopeptidase inhibitor 1 (*TIMPI*) gene seemed to predispose to expression from the inactive X (Anderson and Brown 2005).

3.3.3.3 Variability Between Tissues in Escape from XCI

Tissue-specific escape from XCI was reported for mouse lysine (K)-specific demethylase 5C (*Kdm5c*) (Carrel et al. 1996). Using DNA methylation levels as a means of identifying genes that escape XCI, Cotton et al. (2011) predicted the proportion of genes escaping XCI to range from 9 % in blood to 25 % in neural tissue, although the sample size was small (Cotton et al. 2011). Figure 3.4d shows an example of this variability.

3.3.3.4 Is “Escape” Reactivation?

As a final caveat, when a gene is said to escape XCI, it is not known whether it fails to respond to the initial inactivating signal, or whether it tends to reactivate in a substantial proportion of cells. For mouse genes, expression can be monitored early in development and it was shown for one gene that there is initial silencing followed by reactivation (Lingenfelter et al. 1998). It seems likely that both resistance to inactivation and reactivation contribute to this phenomenon. Distinguishing between these alternatives will be important not only for understanding the mechanisms of XCI, but to assess whether reactivation may have implications for disease predisposition with aging.

3.3.4 Impact of Escapee Genes on Phenotype

Escape from XCI will result in higher gene expression in female cells relative to male cells and will eliminate the mosaicism in female carriers to the extent that the two Xs are equivalently expressed. A growing number of X-linked disease genes have been identified to escape XCI including steroid sulfatase (microsomal) isozyme S (*STS*), *KALI* inhibitor of kappa light polypeptide gene enhancer in B-cells kinase gamma (*IKBKG*), and premature ovarian failure 1B (*POF1B*), and the variabilities discussed above may contribute to the variable outcomes seen in carriers. For PAR genes, dosage will be balanced, and the inheritance pattern distinct from both X and autosomal genes. The presence of a Y homolog might compensate for other escape genes; however, it was shown in mouse brain that several genes with Y homologues displayed divergence of gene regulatory sequences between the X and Y (Xu et al. 2008a, b).

Genes escaping XCI have a pronounced effect on phenotype in X-chromosomal aneuploidies. The majority of 45,X conceptuses are lost prior to birth, presumably due to the deficiency of gene products normally expressed from both the X and the Y.

Gain of an X has less severe phenotypic consequences; however, reduced lifespan and some recurrent features may be attributable to over-expression of genes escaping XCI (reviewed in Yang et al. 2011). Additionally, presence of an additional 5 % of the genome may in itself be detrimental to some cellular processes.

Genes that escape from XCI have the potential to contribute to a sex difference in disease susceptibility. Indeed, expression of genes that are variable in escape between females could contribute to inter-female differences as well. In addition, genes that are expressed from the inactive X will be resistant to acquired somatic mutations in females which could be protective against loss of tumor suppressor gene activity resulting in a male predominance to some cancers. In mice a clever breeding strategy has utilized an *Sry*-deficient Y chromosome and *Sry* transgenics to generate XX and XY female mice and XXSry and XY male mice to separate the impact of the sex chromosomes from that of the *Sry* sex-determining gene. Such mice have been used to demonstrate a role for the XX chromosome complement in various disorders including autoimmune disease (Smith-Bouvier et al. 2008). Interestingly, mice appear to have fewer genes that escape from XCI than humans (Berletch et al. 2011), and therefore this mouse model may not reflect the full extent of sex differences attributable to the X in human disease situations. Another difference between humans and mice is the presence of imprinted genes on the mouse X, which have not yet been identified in humans (Sharp et al. 2011). As males always inherit only a maternal X, imprinted genes have the potential for considerable differences in expression between males and females. A further consideration for the impact of XCI on human disease is that because the inactive X is largely facultative heterochromatin, it can act either as a sink for heterochromatic proteins, or a tank, representing a potential storehouse of such proteins (Blewitt et al. 2005; Juriloff and Harris 2012). With the emerging role of chromatin remodelers in human disease and cancer, the presence of an inactive X may thus have epigenetic impacts that have not yet been discovered. Overall, in considering X-linked disease in females, the parental origin of new mutations, the incomplete nature of dosage compensation, and the variability in XCI skewing between and within females all need to be considered.

References

- Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW (1992) Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 51:1229–1239
- Amos-Landgraf JM, Cottle A, Plenge RM, Friez M, Schwartz CE, Longshore J, Willard HF (2006) X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am J Hum Genet* 79(3):493–499
- Anderson C, Brown CJ (1999) Polymorphic X-chromosome inactivation of the human TIMP1 gene. *Am J Hum Genet* 65:699–708
- Anderson CL, Brown CJ (2005) Epigenetic predisposition to expression of TIMP1 from the human inactive X chromosome. *BMC Genet* 6:48

- Asplund A, Guo Z, Hu X, Wassberg C, Ponten F (2001) Mosaic pattern of maternal and paternal keratinocyte clones in normal human epidermis revealed by analysis of X-chromosome inactivation. *J Invest Dermatol* 117(1):128–131
- Au WY, Lam V, Pang A, Lee WM, Chan JL, Song YQ, Ma ES, Kwong YL (2006) Glucose-6-phosphate dehydrogenase deficiency in female octogenarians, nanogenarians, and centenarians. *J Gerontol A Biol Sci Med Sci* 61(10):1086–1089
- Beever C, Lai BP, Baldry SE, Penaherrera MS, Jiang R, Robinson WP, Brown CJ (2003a) Methylation of ZNF261 as an assay for determining X chromosome inactivation patterns. *Am J Med Genet A* 120A(3):439–441
- Beever CL, Stephenson MD, Penaherrera MS, Jiang RH, Kalousek DK, Hayden M, Field L, Brown CJ, Robinson WP (2003b) Skewed X-chromosome inactivation is associated with trisomy in women ascertained on the basis of recurrent spontaneous abortion or chromosomally abnormal pregnancies. *Am J Hum Genet* 72(2):399–407
- Berleth JB, Yang F, Disteché CM (2010) Escape from X inactivation in mice and humans. *Genome Biol* 11(6):213
- Berleth JB, Yang F, Xu J, Carrel L, Disteché CM (2011) Genes that escape from X inactivation. *Hum Genet* 130(2):237–245
- Bertelsen B, Tumer Z, Ravn K (2011) Three new loci for determining X chromosome inactivation patterns. *J Mol Diagn* 13(5):537–540
- Bianchi I, Lleo A, Gershwin ME, Ivernizzi P (2012) The X chromosome and immune associated genes. *J Autoimmun* 38(2–3):J187–J192
- Bittel DC, Theodoro MF, Kibiryeva N, Fischer W, Talebizadeh Z, Butler MG (2008) Comparison of X-chromosome inactivation patterns in multiple tissues from human females. *J Med Genet* 45(5):309–313
- Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JJ, Arkell R, Whitelaw E (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A* 102(21):7629–7634
- Bolduc V, Chagnon P, Provost S, Dube MP, Belisle C, Gingras M, Mollica L, Busque L (2008) No evidence that skewing of X chromosome inactivation patterns is transmitted to offspring in humans. *J Clin Invest* 118(1):333–341
- Boyd Y, Buckle VJ (1986) Cytogenetic heterogeneity of translocations associated with Duchenne muscular dystrophy. *Clin Genet* 29(2):108–115
- Bretherick KL, Metzger DL, Chanoine JP, Panagiotopoulos C, Watson SK, Lam WL, Fluker MR, Brown CJ, Robinson WP (2007) Skewed X-chromosome inactivation is associated with primary but not secondary ovarian failure. *Am J Med Genet A* 143A(9):945–951
- Brown CJ, Willard HF (1994) The human X inactivation center is not required for maintenance of X inactivation. *Nature* 368:154–156
- Busque L, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, Maragh M, Gilliland DG (1996) Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood* 88:59–65
- Busque L, Paquette Y, Provost S, Roy DC, Levine RL, Mollica L, Gilliland DG (2009) Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. *Blood* 113(15):3472–3474
- Busque L, Zhu J, DeHart D, Griffith B, Willman C, Carroll R, Black PM, Gilliland DG (1994) An expression based clonality assay at the human androgen receptor locus (HUMARA) on the X chromosome. *Nucleic Acids Res* 22:697–698
- Butler MG, Theodoro MF, Bittel DC, Kuipers PJ, Driscoll DJ, Talebizadeh Z (2007) X-chromosome inactivation patterns in females with Prader-Willi syndrome. *Am J Med Genet A* 143(5):469–475
- Carrel L, Hunt PA, Willard HF (1996) Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. *Hum Mol Genet* 5:1361–1366
- Carrel L, Willard HF (1993) The X-linked choroideremia gene escapes X chromosome inactivation. *Am J Hum Genet* 53(3):42A

- Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434(7031):400–404
- Cattanach BM, Pollard CE, Perez JN (1969) Controlling elements in the mouse X-chromosome 1. Interaction with the X-linked genes. *Genet Res* 14:223–235
- Cau M, Addis M, Congiu R, Meloni C, Cao A, Santaniello S, Loi M, Emma F, Zuffardi O, Ciccone R, Sole G, Melis MA (2006) A locus for familial skewed X chromosome inactivation maps to chromosome Xq25 in a family with a female manifesting Lowe syndrome. *J Hum Genet* 51(11):1030–1036
- Cazzola M, May A, Bergamaschi G, Cerani P, Rosti V, Bishop DF (2000) Familial-skewed X-chromosome inactivation as a predisposing factor for late-onset X-linked sideroblastic anemia in carrier females. *Blood* 96(13):4363–4365
- Chaturvedi V, Chu MS, Carrol BM, Brenner BJ, Nickoloff BJ (2002) Estimation of size of clonal unit for keratinocytes in normal human skin. *Arch Pathol Lab Med* 126(4):420–424
- Chelly J, Marlhens F, Le Marec B, Jeanpierre M, Lambert M, Hamard G, Dutrillaux B, Kaplan JC (1986) De novo DNA microdeletion in a girl with Turner syndrome and Duchenne muscular dystrophy. *Hum Genet* 74(2):193–196
- Chen GL, Prchal JT (2007) X-linked clonality testing: interpretation and limitations. *Blood* 110(5):1411–1419
- Chitnis S, Derom C, Vlietinck R, Derom R, Monteiro J, Gregersen PK (1999) X chromosome-inactivation patterns confirm the late timing of monamniotic-MZ twinning. *Am J Hum Genet* 65:570–571
- Christensen K, Kristiansen M, Hagen-Larsen H, Skytthe A, Bathum L, Jeune B, Andersen-Ranberg K, Vaupel JW, Orstavik KH (2000) X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood* 95:2449–2451
- Chung IM, Schwartz SM, Murry CE (1998) Clonal architecture of normal and atherosclerotic aorta: implications for atherogenesis and vascular development. *Am J Pathol* 152(4):913–923
- Clements SE, Mellerio JE, Holden ST, McCauley J, McGrath JA (2009) PORCN gene mutations and the protean nature of focal dermal hypoplasia. *Br J Dermatol* 160(5):1103–1109
- Cotton AM, Lam L, Affleck JG, Wilson IM, Penaherrera MS, McFadden DE, Kobor MS, Lam WL, Robinson WP, Brown CJ (2011) Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum Genet* 130(2):187–201
- D'Esposito M, Ciccodicola A, Gianfrancesco F, Esposito T, Flagiello L, Mazzarella R, Schlesinger D, D'Urso M (1996) A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. *Nat Genet* 13(2):227–229
- Daye AG, Bottani A, Bouchardy I, Fluss J, Antonarakis SE, Haenggeli CA, Morris MA (2007) MECP2 mutant allele in a boy with Rett syndrome and his unaffected heterozygous mother. *Brain Dev* 29(1):47–50
- De Bonis ML, Cerase A, Matarazzo MR, Ferraro M, Strazzullo M, Hansen RS, Chiurazzi P, Neri G, D'Esposito M (2006) Maintenance of X- and Y-inactivation of the pseudoautosomal (PAR2) gene SPRY3 is independent from DNA methylation and associated to multiple layers of epigenetic modifications. *Hum Mol Genet* 15(7):1123–1132
- Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW, Schlesinger F, Davis CA, Reinke VJ, Gingeras TR, Shendure J, Waterston RH, Oliver B, Lieb JD, Distech CM (2011) Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet* 43(12):1179–1185
- Di Nunzio S, Cecconi M, Passerini L, McMurchy AN, Baron U, Turbachova I, Vignola S, Valencic E, Tommasini A, Junker A, Cazzola G, Olek S, Levings MK, Perroni L, Roncarolo MG, Bacchetta R (2009) Wild-type FOXP3 is selectively active in CD4 + CD25(hi) regulatory T cells of healthy female carriers of different FOXP3 mutations. *Blood* 114(19):4138–4141
- Diallo R, Schaefer KL, Poremba C, Shivazi N, Willmann V, Buerger H, Dockhorn-Dworniczak B, Boecker W (2001) Monoclonality in normal epithelium and in hyperplastic and neoplastic lesions of the breast. *J Pathol* 193(1):27–32

- Dobrovolny R, Dvorakova L, Ledvinova J, Magage S, Bultas J, Lubanda JC, Elleder M, Karetova D, Pavlikova M, Hrebicek M (2005) Relationship between X-inactivation and clinical involvement in Fabry heterozygotes. Eleven novel mutations in the alpha-galactosidase A gene in the Czech and Slovak population. *J Mol Med* 83(8):647–654
- Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, Oosterwijk JC, Ober C (2004) Inheritance of most X-linked traits is not dominant or recessive, just X-linked. *Am J Med Genet A* 129A(2):136–143
- Durmaz E, Zou M, Al-Rijjal RA, Baitei EY, Hammami S, Bircan I, Akcurin S, Meyer B, Shi Y (2013) Novel and de novo PHEX mutations in patients with hypophosphatemic rickets. *Bone* 52(1):286–289
- Francke U, Felsenstein J, Gartler SM, Migeon BR, Dancis J, Seegmiller JE, Bakay F, Nyhan WL (1976) The occurrence of new mutants in the X-linked recessive Lesch-Nyhan disease. *Am J Hum Genet* 28:123–137
- Gale RE, Linch DC (1998) Clonality studies in acute myeloid leukemia. *Leukemia* 12(2):117–120
- Gale RE, Wheadon H, Boulous P, Linch DC (1994) Tissue specificity of X-chromosome inactivation patterns. *Blood* 83:2899–2905
- Gentilini D, Castaldi D, Mari D, Monti D, Franceschi C, Di Blasio AM, Vitale G (2012) Age-dependent skewing of X chromosome inactivation appears delayed in centenarians' offspring. Is there a role for allelic imbalance in healthy aging and longevity? *Aging Cell* 11(2):277–283
- Gibbons RJ, Suthers GK, Wilkie AO, Buckle VJ, Higgs DR (1992) X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *Am J Hum Genet* 51:1136–1149
- Green PM, Saad S, Lewis CM, Giannelli F (1999) Mutation rates in humans. I. Overall and sex-specific rates obtained from a population study of hemophilia B. *Am J Hum Genet* 65(6):1572–1579
- Grimm T, Meng G, Liechti-Gallati S, Bettecken T, Muller CR, Muller B (1994) On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. *J Med Genet* 31(3):183–186
- Hatakeyama C, Anderson CL, Beever CL, Penaherrera MS, Brown CJ, Robinson WP (2004) The dynamics of X-inactivation skewing as women age. *Clin Genet* 66(4):327–332
- Hogge WA, Prosen TL, Lanasa MC, Huber HA, Reeves MF (2007) Recurrent spontaneous abortion and skewed X-inactivation: is there an association? *Am J Obstet Gynecol* 196(4):384, e381–386
- Hopkins II, Humphrey I, Keith CG, Susman M, Webb GC, Turner EK (1979) The Aicardi syndrome in a 47, XXY male. *Aust Paediatr J* 15(4):278–280
- Jobanputra V, Levy B, Kinney A, Brown S, Shirazi M, Yu C, Kline J, Warburton D (2012) Copy number changes on the X chromosome in women with and without highly skewed X-chromosome inactivation. *Cytogenet Genome Res* 136(4):264–269
- Jovanovic L, Delahunt B, McIver B, Eberhardt NL, Grebe SK (2003) Thyroid gland clonality revisited: the embryonal patch size of the normal human thyroid gland is very large, suggesting X-chromosome inactivation tumor clonality studies of thyroid tumors have to be interpreted with caution. *J Clin Endocrinol Metab* 88(7):3284–3291
- Juriloff DM, Harris MJ (2012) Hypothesis: The female excess in cranial neural tube defects reflects an epigenetic drag of the inactivating x chromosome on the molecular mechanisms of neural fold elevation. *Birth Defects Res A Clin Mol Teratol* 94(10):849–855
- Knudsen GP, Pedersen J, Klingenberg O, Lygren I, Orstavik KH (2007) Increased skewing of X chromosome inactivation with age in both blood and buccal cells. *Cytogenet Genome Res* 116(1–2):24–28
- Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WS, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT,

- Thorsteinsdottir U, Stefansson K (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475
- Kristiansen M, Knudsen GP, Bathum L, Naumova AK, Sorensen TI, Brix TH, Svendsen AJ, Christensen K, Kyvik KO, Orstavik KH (2005) Twin study of genetic and aging effects on X chromosome inactivation. *Eur J Hum Genet* 13(5):599–606
- Kubota T, Nonoyama S, Tonoki H, Masuno M, Imaizumi K, Kojima M, Wakui K, Shimadzu M, Fukushima Y (1999) A new assay for the analysis of X-chromosome inactivation based on methylation-specific PCR. *Hum Genet* 104(1):49–55
- Kucera KS, Reddy TE, Pauli F, Gertz J, Logan JE, Myers RM, Willard HF (2011) Allele-specific distribution of RNA polymerase II on female X chromosomes. *Hum Mol Genet* 20(20):3964–3973
- Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. *Science* 286:964–967
- Lanasa MC, Hogge WA, Kubik C, Blancato J, Hoffman EP (1999) Highly skewed X-chromosome inactivation is associated with idiopathic recurrent spontaneous abortion. *Am J Hum Genet* 65(1):252–254
- Lau AW, Brown CJ, Penaherrera M, Langlois S, Kalousek DK, Robinson WP (1997) Skewed X-chromosome inactivation is common in fetuses or newborns associated with confined placental mosaicism. *Am J Hum Genet* 61:1353–1361
- Leppig KA, Distech CM (2001) Ring X and other structural X chromosome abnormalities: X inactivation and phenotype. *Semin Reprod Med* 19(2):147–157. doi:10.1055/s-2001-15395
- Leuer M, Oldenburg J, Laverne JM, Ludwig M, Fregin A, Eigel A, Ljung R, Goodeve A, Peake I, Olek K (2001) Somatic mosaicism in hemophilia A: a fairly common event. *Am J Hum Genet* 69(1):75–87
- Lingenfelter PA, Adler DA, Poslinski D, Thomas S, Elliot RW, Chapman VM, Distech CM (1998) Escape from X inactivation of Smcx is preceded by silencing during mouse development. *Nat Genet* 18:212–213
- Livernois AM, Graves JA, Waters PD (2012) The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity (Edinb)* 108(1):50–58
- Lyon MF (1962) Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* 14:135–145
- Migeon BR (2006) The role of X inactivation and cellular mosaicism in women's health and sex-specific diseases. *JAMA* 295(12):1428–1433
- Migeon BR, Shapiro LJ, Norum RA, Mohandas T, Axelman J, Dabora RL (1982) Differential expression of steroid sulphatase locus on active and inactive human X chromosome. *Nature* 299:838–840
- Minks J, Robinson WP, Brown CJ (2008) A skewed view of X chromosome inactivation. *J Clin Invest* 118(1):20–23
- Monteiro J, Derom C, Vlietinck R, Kohn N, Lesser M, Gregersen PK (1998) Commitment to X inactivation precedes the twinning event in monochorionic MZ twins. *Am J Hum Genet* 63:339–346
- Mutter GL, Boynton KA, Faquin WC, Ruiz RE, Jovanovic AS (1996) Allelotype mapping of unstable microsatellites establishes direct lineage continuity between endometrial precancers and cancer. *Cancer Res* 56(19):4483–4486
- Naumova AK, Olien L, Bird LM, Smith M, Verner AE, Leppert M, Morgan K, Sapienza C (1998) Genetic mapping of X-linked loci involved in skewing of X chromosome inactivation in the human. *Eur J Hum Genet* 6:552–562
- Nguyen DK, Distech CM (2006) Dosage compensation of the active X chromosome in mammals. *Nat Genet* 38(1):47–53
- Novelli M, Cossu A, Ouakrif D, Quaglia A, Lakhani S, Poulsom R, Sasiene P, Carta P, Contini M, Pasca A, Palmieri G, Bodmer W, Tanda F, Wright N (2003) X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. *Proc Natl Acad Sci U S A* 100(6):3311–3314

- Orstavik KH (2009) X chromosome inactivation in clinical practice. *Hum Genet* 126(3):363–373
- Ozbalkan Z, Bagislar S, Kiraz S, Akyerli CB, Ozer HT, Yavuz S, Birlik AM, Calguneri M, Ozcelik T (2005) Skewed X chromosome inactivation in blood cells of women with scleroderma. *Arthritis Rheum* 52(5):1564–1570
- Ozcelik T, Uz E, Akyerli CB, Bagislar S, Mustafa CA, Gursay A, Akarsu N, Toruner G, Kamel N, Gullu S (2006) Evidence from autoimmune thyroiditis of skewed X-chromosome inactivation in female predisposition to autoimmunity. *Eur J Hum Genet* 14(6):791–797
- Pegoraro E, Whitaker J, Mowery-Rushton P, Surti U, Lanasa M, Hoffman E (1998) Familial skewed X inactivation: a molecular trait associated with high spontaneous-abortion rate maps to Xq28. *Am J Hum Genet* 61:160–170
- Penaherrera MS, Jiang R, Avila L, Yuen RK, Brown CJ, Robinson WP (2012) Patterns of placental development evaluated by X chromosome inactivation profiling provide a basis to evaluate the origin of epigenetic variation. *Hum Reprod* 27(6):1745–1753
- Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. *Nature* 379:131–137
- Pereira LV, Zata M (1999) Screening of the C43G mutation in the promoter region of the XIST gene in females with highly skewed X-chromosome inactivation. *Am J Med Genet* 87:86–87
- Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GA (2012) Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. *Proc Natl Acad Sci U S A* 109(14):5346–5351
- Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, Winter RM, Willard HF (1997) A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. *Nat Genet* 17:353–356
- Puck JM, Willard HF (1998) X inactivation in females with X-linked disease. *New Engl J Med* 338:325–328
- Pugacheva EM, Tiwari VK, Abdullaev Z, Vostrov AA, Flanagan PT, Quitschke WW, Loukinov DI, Ohlsson R, Lobanenkov VV (2005) Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Hum Mol Genet* 14(7):953–965
- Renault NKE, Renault MP, Copeland E, Howell RE, Greer WL (2011) Familial skewed X-chromosome inactivation linked to a component of the cohesin complex, SA2. *J Hum Genet* 56(5):390–397
- Richards CS, Watkins SC, Hoffman EP, Schneider NR, Milsark IW, Katz KS, Cook JD, Kunkel LM, Cortada JM (1990) Skewed X inactivation in a female MZ twin results in Duchenne muscular dystrophy. *Am J Hum Genet* 46(4):672–681
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurler ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RI, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnini D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen Z, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Cobley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, Deshazo D, Dhami P, Ding Y, Dinh H, Dodsworth S, Draper H, Dugan-Rocha S, Dunham A, Dunn M, Durbin KJ, Dutta I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill R, Glockner G, Gregory SG, Gribble S, Griffiths C, Grocock R, Gu Y, Gwilliam R, Hamilton C, Hart EA, Hawes A, Heath PD, Heitmann K, Hennig S, Hernandez J, Hinzmann B, Ho S, Hoffmann M, Howden PJ, Huckle EJ, Hume J, Hunt PJ,

- Hunt AR, Isherwood J, Jacob L, Johnson D, Jones S, de Jong PJ, Joseph SS, Keenan S, Kelly S, Kershaw JK, Khan Z, Kioschis P, Klages S, Knights AJ, Kosiura A, Kovar-Smith C, Laird GK, Langford C, Lawlor S, Leversha M, Lewis L, Liu W, Lloyd C, Lloyd DM, Loulseghed H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma J, Maheshwari M, Matthews LH, McDowall J, McLaren S, McMurray A, Meidl P, Meitinger T, Milne S, Miner G, Mistry SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsiek G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Pelan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridler KA, Schlessinger D, Schueler MG, Sehra HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Sotheran EC, Steingruber HE, Steward CA, Storey R, Swann RM, Swarbreck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevanion S, Tromans AC, d'Urso M, Verduzco D, Villasana D, Waldron L, Wall M, Wang Q, Warren J, Warry GL, Wei X, West A, Whitehead SL, Whiteley MN, Wilkinson JE, Willey DL, Williams G, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang J, Zhou J, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poustka A, Rosenthal A, Lehrach H, Meindl A, Minx PJ, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM, Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulston JE, Durbin R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DR (2005) The DNA sequence of the human X chromosome. *Nature* 434(7031):325–337
- Ryan SG, Chance PF, Zou CH, Spinner NB, Golden JA, Smietana S (1997) Epilepsy and mental retardation limited to females: an X-linked dominant disorder with male sparing. *Nat Genet* 17(1):92–95
- Salsano E, Tabano S, Sirchia SM, Colapietro P, Castellotti B, Gellera C, Rimoldi M, Pensato V, Mariotti C, Pareyson D, Miozzo M, Uziel G (2012) Preferential expression of mutant ABCD1 allele is common in adrenoleukodystrophy female carriers but unrelated to clinical symptoms. *Orphanet J Rare Dis* 7:10
- Sato K, Uehara S, Hashiyada M, Nabeshima H, Sugawara J, Terada Y, Yaegashi N, Okamura K (2004) Genetic significance of skewed X-chromosome inactivation in premature ovarian failure. *Am J Med Genet A* 130A(3):240–244
- Schluth C, Cossee M, Girard-Lemaire F, Carelle N, Dollfus H, Jeandidier E, Flori E (2007) Phenotype in X chromosome rearrangements: pitfalls of X inactivation study. *Pathol Biol* 55(1):29–36
- Sharp A, Robinson D, Jacobs P (2000) Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Hum Genet* 107(4):343–349
- Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, Antonarakis SE (2011) DNA methylation profiles of human active and inactive X chromosomes. *Genome Res* 21(10):1592–1600. doi:10.1101/gr.112680.110
- Smeets EE, Pelc K, Dan B (2012) Rett syndrome. *Mol Syndromol* 2(3–5):113–127
- Smith-Bouvier DL, Divekar AA, Sasidhar M, Du S, Tiwari-Woodruff SK, King JK, Arnold AP, Singh RR, Voskuhl RR (2008) A role for sex chromosome complement in the female bias in autoimmune disease. *J Exp Med* 205(5):1099–1108
- Stephenson MD, Awartani KA, Robinson WP (2002) Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. *Hum Reprod* 17(2):446–451
- Su MT, Lin SH, Chen YC (2011) Association of sex hormone receptor gene polymorphisms with recurrent pregnancy loss: a systematic review and meta-analysis. *Fertil Steril* 96(6):1435–1444
- Swierczek SI, Piterkova L, Jelinek J, Agarwal N, Hammoud S, Wilson A, Hickman K, Parker CJ, Cairns B, Prchal JT (2012) Methylation of AR locus does not always reflect X chromosome inactivation state. *Blood* 119(13):e100–e109
- Tanaka M, Kyo S, Kanaya T, Yatabe N, Nakamura M, Maida Y, Okabe M, Inoue M (2003) Evidence of the monoclonal composition of human endometrial epithelial glands and mosaic pattern of clonal distribution in luminal epithelium. *Am J Pathol* 163(1):295–301

- Tomkins DJ, McDonald HL, Farrell SA, Brown CJ (2002) Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay. *Eur J Hum Genet* 10(1):44–51
- Toniolo D, Rizzolio F (2007) X chromosome and ovarian failure. *Semin Reprod Med* 25(4):264–271
- Twigg SR, Kan R, Babbs C, Bochukova EG, Robertson SP, Wall SA, Morriss-Kay GM, Wilkie AO (2004) Mutations of ephrin-B1 (EFNB1), a marker of tissue boundary formation, cause craniofrontonasal syndrome. *Proc Natl Acad Sci U S A* 101(23):8652–8657
- Uehara S, Hashiyada M, Sato K, Sato Y, Fujimori K, Okamura K (2001) Preferential X-chromosome inactivation in women with idiopathic recurrent pregnancy loss. *Fertil Steril* 76(5):908–914
- Uz E, Mustafa C, Topaloglu R, Bilginer Y, Dursun A, Kasapcopur O, Ozen S, Bakkaloglu A, Ozelik T (2009) Increased frequency of extremely skewed X chromosome inactivation in juvenile idiopathic arthritis. *Arthritis Rheum* 60(11):3410–3412
- van Dijk JP, Heuver LH, van der Reijden BA, Raymakers RA, de Witte T, Jansen JH (2002) A novel, essential control for clonality analysis with human androgen receptor gene polymerase chain reaction. *Am J Pathol* 161(3):807–812
- Van Vlierberghe P, Patel J, Abdel-Wahab O, Lobry C, Hedvat CV, Balbin M, Nicolas C, Payer AR, Fernandez HF, Tallman MS, Paietta E, Melnick A, Vandenberghe P, Speleman F, Aifantis I, Cools J, Levine R, Ferrando A (2011) PHF6 mutations in adult acute myeloid leukemia. *Leukemia* 25(1):130–134
- Vickers MA, McLeod E, Spector TD, Wilson IJ (2001) Assessment of mechanism of acquired skewed X inactivation by analysis of twins. *Blood* 97(5):1274–1281
- Wang X, Reid Sutton V, Omar Peraza-Llanes J, Yu Z, Rosetta R, Kou YC, Eble TN, Patel A, Thaller C, Fang P, Van den Veyver IB (2007) Mutations in X-linked PORCN, a putative regulator of Wnt signaling, cause focal dermal hypoplasia. *Nat Genet* 39(7):836–838
- Wang Y, Busin R, Reeves C, Bezman L, Raymond G, Toomer CJ, Watkins PA, Snowden A, Moser A, Naidu S, Bibat G, Hewson S, Tam K, Clarke JT, Charnas L, Stetten G, Karczeski B, Cutting G, Steinberg S (2011) X-linked adrenoleukodystrophy: ABCD1 de novo mutations and mosaicism. *Mol Genet Metab* 104(1–2):160–166
- Warburton D, Kline J, Kinney A, Yu CY, Levin B, Brown S (2009) Skewed X chromosome inactivation and trisomic spontaneous abortion: no association. *Am J Hum Genet* 85(2):179–193
- Wieacker P, Wieland I (2005) Clinical and genetic aspects of craniofrontonasal syndrome: towards resolving a genetic paradox. *Mol Genet Metab* 86(1–2):110–116
- Xu J, Deng X, Disteche CM (2008a) Sex-specific expression of the X-linked histone demethylase gene *Jarid1c* in brain. *PLoS One* 3(7):e2553
- Xu J, Deng X, Watkins R, Disteche CM (2008b) Sex-specific differences in expression of histone demethylases *Utx* and *Uty* in mouse brain and neurons. *J Neurosci* 28(17):4521–4527
- Yang C, Chapman AG, Kelsey AD, Minks J, Cotton AM, Brown CJ (2011) X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum Genet* 130(2):175–185
- Yang F, Babak T, Shendure J, Disteche CM (2010) Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* 20(5):614–622
- Yasukochi Y, Maruyama O, Mahajan MC, Padden C, Euskirchen GM, Schulz V, Hirakawa H, Kuhara S, Pan XH, Newburger PE, Snyder M, Weissman SM (2010) X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. *Proc Natl Acad Sci U S A* 107(8):3704–3709
- Zhang J, Bao X, Cao G, Jiang S, Zhu X, Lu H, Jia L, Pan H, Fehr S, Davis M, Leonard H, Ravine D, Wu X (2011) What does the nature of the MECP2 mutation tell us about parental origin and recurrence risk in Rett syndrome? *Clin Genet* 82(6):526–533

Chapter 4

***Cis-* and *Trans*-Effects Underlying Polar Overdominance at the Callipyge Locus**

Michel Georges, Haruko Takeda, Huijun Cheng, Xu Xuewen,
Tracy Hadfield-Shay, Noelle Cockett, and Carole Charlier

Abstract The callipyge phenotype is a muscular hypertrophy of sheep that is characterized by a unique mode of inheritance, referred to as polar overdominance, in which only heterozygous individuals inheriting the *CLPG* mutation from their sire express the phenotype. We herein report recent advances towards understanding the molecular mechanisms underlying polar overdominance. They involve an interplay between *cis*- and *trans*-effects of the *CLPG* mutation (Fig. 4.1).

The callipyge phenotype is a muscular hypertrophy of sheep that is characterized by a unique mode of inheritance, referred to as polar overdominance, in which only heterozygous individuals inheriting the *CLPG* mutation from their sire express the phenotype. We herein report recent advances towards understanding the molecular mechanisms underlying polar overdominance. They involve an interplay between *cis*- and *trans*-effects of the *CLPG* mutation (Fig. 4.1).

M. Georges (✉) • H. Takeda • H. Cheng • X. Xuewen • C. Charlier
Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège
(B34), 1 Avenue de l'Hôpital, 4000 Liège, Belgium
e-mail: michel.georges@ulg.ac.be; htakeda@ulg.ac.be; huijun.cheng@student.ulg.ac.be;
xuewen_xu@mail.hzau.edu.cn; carole.charlier@ulg.ac.be

T. Hadfield-Shay • N. Cockett
Department of Animal, Dairy, and Veterinary Sciences, Utah State University,
4700 Old Main Hill, Logan, UT 84322-4700, USA
e-mail: tracy.hadfield.shay@cc.usu.edu; noelle.cockett@usu.edu

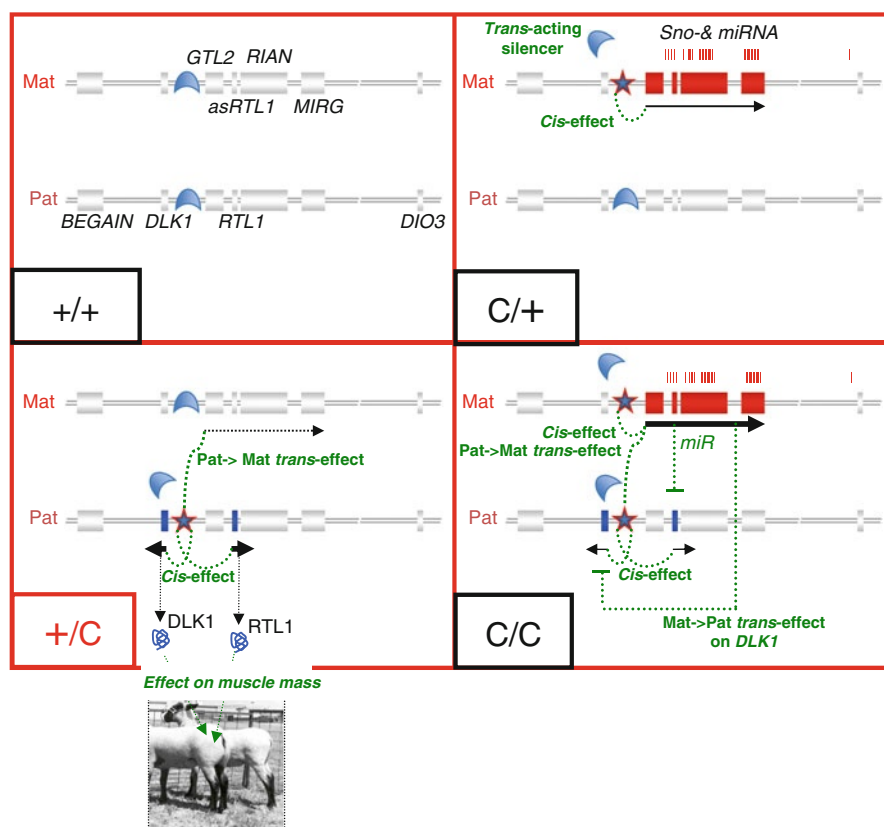


Fig. 4.1 Working model for polar overdominance at the ovine callipyge locus. The *boxes* illustrate the expression profile of the *BEGAIN-DIO3* domain in skeletal muscle according to *CLPG* genotype. ★: *CLPG* mutation

4.1 The Callipyge Phenotype

“Callipyge” (from Greek *calli*:- beautiful and *pyge*: buttocks) is the name given to a heritable muscular hypertrophy of sheep (Cockett et al. 1994). It was first observed in a farm in Oklahoma in the 1980s, where it reportedly affected ~10–15 % of the offspring of a ram named “Solid Gold.” When slaughtered at equal live weight as controls, individual muscle of callipyge lambs weight on average 38 % more. The hypertrophy is more pronounced for the muscles of the pelvic limb (average: 42 %; range: 12–58 %) and torso (average: 50 %; range: 51–39 %) than for the thoracic limb (average: 14 %; range: 6–22 %). The muscular hypertrophy is sufficiently pronounced to allow experienced observers to unambiguously recognize animals with the condition. Callipyge animals do not grow faster than controls, but are more feed efficient. The meat of callipyge animals is leaner yet tougher than that of controls, and this has hampered their widespread commercial use. The muscular

hypertrophy was shown to be due to an increase in the proportion and diameter of fast twitch glycolytic muscle fibers (Koohmaraie et al. 1995; Carpenter et al. 1996; Jackson et al. 1997a–c; Freking et al. 1998a, b, 1999; Freking and Leymaster 2006).

4.2 Polar Overdominance

When mated to wild-type ewes, callipyge rams descendent of Solid Gold produced ~50 % callipyge offspring, irrespective of gender. In these crosses, the callipyge phenotype thus appeared to be autosomal dominant. A linkage scan conducted under this model indeed identified a locus on distal chromosome 18 that fully accounted for the segregation of the trait in these pedigrees. The locus and corresponding mutation were labeled *CLPG* (Cockett et al. 1994). Surprisingly, the reciprocal mating (namely between callipyge ewes and wild-type rams) produced only wild-type offspring, despite transmission of the *CLPG* haplotype to half of the offspring, as expected. This nonequivalence of reciprocal crosses suggested that the *CLPG* locus might be subject to parental imprinting with epigenetic silencing of the maternal (i.e., transmitted by the mother) allele. In agreement with the imprinting hypothesis, wild-type sons having received the *CLPG* haplotype from their callipyge dam (genotype *CLPG^{Mat}/+^{Pat}*) produced 50 % callipyge offspring when mated to unrelated wild-type ewes (genotype *+/+*) (Cockett et al. 1996).

Parental imprinting, however, did not account for the observation that only ~25 % of offspring from crosses between callipyge sires and dams (genotype: *+^{Mat}/CLPG^{Pat}*) expressed the phenotype. Assuming a silenced maternal *CLPG* allele, *+^{Mat}/CLPG^{Pat}* as well as *CLPG/CLPG* offspring were expected to be callipyge. Yet, if *+^{Mat}/CLPG^{Pat}* offspring were indeed found to be callipyge in these matings, their *CLPG/CLPG* (i.e., homozygous for the *CLPG* haplotype) sibs were not. We have called this unusual non-Mendelian inheritance pattern (in which only heterozygous individuals inheriting the mutation from a sex-specified parent have a superior phenotype) “polar overdominance” (Cockett et al. 1996). It was subsequently confirmed by others (f.i. Freking et al. 1998a).

The polar overdominance model predicted that matings between wild-type *CLPG/CLPG* rams and wild-type *+/+* ewes should yield 100 % callipyge offspring and this was indeed found to be the case (Cockett et al. 1996).

4.3 The *BEGAIN-DIO3* Imprinted Domain

To decipher the molecular mechanism underlying polar overdominance, the *CLPG* mutation was first fine-mapped to a ~4 Mb and then to a ~400 kb chromosome interval on distal chromosome 18 (Fahrenkrug et al. 2000; Shay et al. 2001; Berghmans et al. 2001). The interval was shown to encompass a novel imprinted domain that was just being discovered (f.i. da Rocha et al. 2008).

The *BEGAIN-DIO3* domain is now known to span ~1 Mb in eutherians and to encompass eight genes subject to parental imprinting in placental mammals but not in marsupials and monotremes (Charlier et al. 2001a; Paulsen et al. 2001; Edwards et al. 2008). Strikingly, the four genes that are preferentially expressed from the padumnal (i.e., transmitted by the father) allele are protein-coding genes (*BEGAIN*, *DLK1*, *RTL1* (also known as *PEG11*), and *DIO3*), while the four that are preferentially expressed from the madumnal allele are long, noncoding RNA genes (lncRNAs) (*GTL2* (also known as *MEG3*), *RTL1-AS* (also known as *PEG11-AS*), *RIAN* (also known as *MEG8*), and *MIRG* (also known as *MEG9*)).

BEGAIN (brain-enriched guanylate kinase-associated protein) was originally isolated from a rat brain cDNA library and shown to encode a protein that binds to the guanylate kinase domain of PSD95/SAP90, a scaffolding protein at the post-synaptic cell membrane (Deguchi et al. 1998). We have shown in sheep that *BEGAIN* is actually very broadly expressed, easily detectable in multiple tissues both pre- and postnatally. Alternative usage of two promoters (A and B) and splicing generates multiple isoforms. Transcripts initiated at the B promoter exhibit clear imprinting with predominant expression from the padumnal allele in most tissues (but not liver). Transcripts initiated at the A promoter are biallelically expressed in most tissues with, however, predominance of the padumnal allele in kidney (Smit et al. 2005). Tierling et al. (2009) subsequently showed that *BEGAIN* is likewise imprinted in the mouse in a tissue- and promoter-specific manner and controlled by the IG-DMR located upstream of *GTL2* (see hereafter).

DLK1 (delta-like homologue 1; also known as preadipocyte factor-1 (*PREF1*) or fetal antigen (*FAI*)) encodes a so-called noncanonical Notch ligand characterized by six EGF-like domains, a trans-membrane domain and a short intracellular tail. It misses the N-terminal Delta-Serrate-LAG-2 (DSL) domain, involved in receptor binding, shared by all canonical Notch ligands. Full-length *DLK1* possesses a juxtamembrane TACE(ADAM17)-mediated cleavage site, which—upon proteolysis—releases a soluble form of *DLK1*. Alternative splice-variants skip this site, encoding membrane-bound *DLK1* isoforms (f.i. Falix et al. 2012). *DLK1* is expressed in a wide range of tissues (including placenta, liver, adipose, skeletal muscle, lung, vertebrae, pituitary and adrenal glands) before birth, while its expression in adults is limited to the pituitary and adrenal glands, pancreas, central nervous system, testes, prostate, and ovaries. *DLK1* has been shown to play essential roles in adipogenesis, hematopoiesis, neurogenesis, development of skeletal muscle, liver, lung, pancreas, and pituitary gland and adaptation to independent life (f.i. Charalambous et al. 2012; Falix et al. 2012; Mirshekar-Syahkal et al. 2013; Waddell et al. 2010). *DLK1* appears up-regulated in specific tumors including neuroblastoma, hepatoblastoma, and Wilms tumors (f.i. Falix et al. 2012). *DLK1* is thought to directly interact thereby inhibiting Notch receptors (f.i. Bray et al. 2008). Soluble *DLK1* secreted by niche astrocytes was recently shown to stimulate proliferation of neural stem cells of the subventricular zone expressing membrane-bound *DLK1*, suggesting signaling by direct interaction between the two isoforms (Ferrón et al. 2011).

RTL1 (*Retroposon like 1*) is one of nine mammalian genes derived from ancestral sushi-ichi LTR elements of the Ty3/gypsy family that were most likely acquired by horizontal gene transfer (f.i. Youngson et al. 2005). *RTL1* has lost its LTR sequences, yet has maintained a highly conserved, intron-less ~4 kb open reading frame homologous to the gag and pol genes of the retroelement from which it derives, strongly suggesting eutherian-specific “exaptation.” Knockout and overexpression experiments indicate that *RTL1* plays an essential role in the development of the placenta, an eutherian-specific organ (Sekita et al. 2008). *RTL1* has been found to be heavily methylated on both alleles in all examined tissues.

DIO3 encodes the type 3 deiodinase (D3) that inactivates both T3 and T4 by 5-deiodination of the inner ring and acts locally to reduce TH availability (Tsai et al. 2002). Placental expression of *DIO3* protects the fetus against thyrotoxicosis due to high levels of thyroid hormone in the maternal circulation.

The lncRNA *GTL2* (*Gene trap locus 2*) encompasses at least 12 exons and generates a panoply of alternatively spliced transcripts (Schuster-Gossler et al. 1998). Surprisingly, although the exon–intron structure of *GTL2* appears to be conserved amongst eutherians, the exonic sequences per se are not more conserved than the intronic sequences, raising questions about the nature of its function (f.i. Charlier et al. 2001a). *GTL2* has recently been shown to directly bind the polycomb repressive complex 2 (PRC2), suggesting that it may guide PRC2 to impose repressive chromatin marks at specific chromosomal target regions. *GTL2* knock-down in embryonic stem cells causes a twofold increase in *DLK1* transcript levels and reduction in H3K27 trimethylation at the *DLK1* promoter, suggesting that one of the functions of *GTL2* might be to *cis*-regulate the madumal *DLK1* allele by blocking its transcription (Zhao et al. 2010).

As its name implies, *RTL1-AS* is a lnc RNA that is antisense to the intron-less *RTL1* gene. *RTL1-AS* forms six hairpin loops that are recognized by the DROSHA-DGCR8 microprocessor complex to generate pre-miRNAs that will be further processed to generate a quadrille of miRNAs (Seitz et al. 2003; Davis et al. 2005). Being perfectly complementary to *RTL1* over their entire length, this set of miRNAs has indeed been shown to mediate slicing of *RTL1* (assumed to be mediated by AGO2), a mode-of-action of miRNA which, although common in plants, is exceptional in animals (Davis et al. 2005). The strand-specific acquisition of miRNA precursors in a maternally expressed transcript that is antisense to a paternally expressed gene encoding a protein that is essential for placental development is a remarkable feat, thought to be driven by the evolutionary forces underlying the kinship theory of parental imprinting (Haig 2000).

The most striking distinctive feature of the *RIAN* lncRNA gene is the fact that it encodes more than 40 small RNAs resembling small nucleolar RNAs (snoRNAs) of the C/D type falling in two main clusters (*SNORD113* and *SNORD114*) (Cavaillé et al. 2002). The primary known function of C/D snoRNAs is the 2'-O-methylation of rRNAs, small nuclear RNAs, and tRNAs. However, none of the C/D snoRNAs from the *BEGAIN-DIO3* domain exhibits significant sequence complementarity to any one of these putative targets. Also, the C/D snoRNAs do not correspond to peaks of evolutionary conservation in the *RIAN* gene (Caiment et al. 2010).

Twelve of the C/D snoRNAs have been shown to be precursors of miRNAs in sheep (Caiment et al. 2010) and eight to be precursors of Piwi-interacting RNAs (piRNAs) expressed in testis (Girard et al. 2006). C/D snoRNAs from the *BEGAIN-DIO3* domain have been shown to be overexpressed in some patients with acute myeloid leukemia (AML), and their in vitro overexpression to induce cell proliferation (reviewed in Girardot et al. 2012).

MIRG is the pri-miRNA of a cluster of ~50 miRNAs (Seitz et al. 2004; Glazov et al. 2008; Caiment et al. 2010); reviewed in Girardot et al. 2012). The positions of the *MIRG* miRNA coincide with peaks of evolutionary conservation testifying for an essential, sequence-dependent function. A small subset of the *MIRG* miRNA undergoes A-to-I editing affecting the seed, which might modulate the target spectrum of the corresponding miRNAs (Kawahara et al. 2007; Caiment et al. 2010). The targets of the *MIRG* miRNA remains essentially unknown, yet seed-centered target prediction reveals an enrichment of regulators of the gene circuitry operating at the transcriptional, translational, and posttranslational level, primarily in the nervous system (f.i. Caiment et al. 2010). Experimental evidence supports a role in the regulation of synaptic development and function. *MIRG* miRNA are found overexpressed in several cancers, and several of them have been linked to stem cells and pluripotency (reviewed in Girardot et al. 2012).

Monoallelic, imprinted expression of all genes in the *BEGAIN-DIO3* domain is controlled by an intergenic germ line differentially methylated region (IGDMR): an ~8 kb sequence associated with a CpG island located ~15 kb upstream of *GTL2* (Takada et al. 2002). The IGDMR is one of three imprinting control regions (ICR) that acquire methylation in the male (rather than female) germ line (in addition to the ICR of the *Igf2-H19* and *Rasgrf1* loci), while remaining unmethylated on maternal transmission (Ferguson-Smith 2011). Deletion of the *BEGAIN-DIO3* IGDMR on the madumal chromosome causes it to behave as a padumal chromosome in the embryo: expression of the protein-encoding *BEGAIN*, *DLK1*, *RTL1*, and *DIO3* genes, and silencing of the noncoding *GTL2*, *RTL1AS*, *RIAN*, and *MIRG* genes. Deletion of the IGDMR on the padumal chromosome does not affect its behavior (Lin et al. 2003). Monoallelic expression is accompanied by somatic methylation of secondary DMRs in at least the *DLK1* and *GTL2* genes (Takada et al. 2002; reviewed in da Rocha et al. 2008). It is noteworthy that the effect of madumal IGDMR deletion differs in placenta, causing expression of the protein-encoding genes (normally only expressed from the padumal allele) without silencing of the noncoding RNA genes (Lin et al. 2007).

4.4 The *CLPG* Mutation

A reference sequence for most of the ovine *CLPG* domain was first generated by sequencing five BAC clones that jointly covered ~570 kb ranging from ~15 kb upstream of *BEGAIN* to ~80 kb downstream of *MIRG* (Charlier et al. 2001a; Smit et al. 2005). Two teams independently used a large-scale PCR-based strategy to

sequence the domain for a chromosome known to carry the *CLPG* mutation. The two teams applied a similar strategy to select the animals to resequence: large numbers of callipyge animals, known to be of $+^{Mat}/CLPG^{Pat}$ genotype, were screened in order to identify individuals that would be autozygous for the *CLPG* marker haplotype. Resequencing such individuals would simultaneously provide sequence information of the haplotype carrying the *CLPG* mutation as well as of a closely related (identical-by-descent) wild-type haplotype. The prediction was that both haplotypes would differ at a very small number of sites, including the *CLPG* mutation. As a matter of fact, both efforts identified the same unique variant differentiating the *CLPG* and $+$ allele: an A to G transition located in the ~90 kb *DLK1-GTL2* intergenic region, 32.7 kb upstream of the *GTL2* transcription start site. The A to G substitution was shown to affect the third position of a dodecamer motif that is virtually perfectly conserved amongst eutherians. In sheep, the G allele was exclusively observed amongst descendants of Solid Gold. Intriguingly, the available elephant sequence has the same G residue at the corresponding position (Freking et al. 2002; Smit et al. 2003).

Solid Gold, the alleged founder of the callipyge flock, was genotyped for the *CLPG* mutation using DNA extracted from leucocytes. He was found to be heterozygous A/G. Yet, the allelic ratio departed markedly from the expected 1:1, being closer to 0.8(A):0.2(G). Genotyping a panel of microsatellites excluded leucochimerism, leaving mosaicism as most likely explanation. It indicated that the corresponding A to G transition occurred during early development of Solid Gold, proving beyond any reasonable doubt that it is indeed the *CLPG* mutation. The fact that only 15 % of Solid Gold's offspring were callipyge is probably reflecting the fact that he was germ line mosaic as well (Smit et al. 2003).

4.5 *Cis*-Effects of the *CLPG* Mutation

The strong sequence conservation of the dodecamer motif suggested that it might act as *cis*-acting regulatory element that might be perturbed by the *CLPG* mutation. Indeed, it was observed that on paternal transmission of the *CLPG* mutation, *DLK1* and *RTL1* transcript levels remained elevated in postnatal skeletal muscle at ages where these genes are normally nearly completely silenced, and that all corresponding transcripts originated from the paternal allele. Accordingly, on maternal transmission of the *CLPG* mutation, levels of *GTL2*, *RTL1-AS*, *RIAN*, *MIRG*, and associated sno- and miRNAs remained high in the same tissue and at the same developmental stage, all corresponding transcripts originating from the maternal allele. Transcript levels of *BEGAIN* and *DIO3* were unaffected by the *CLPG* mutation whether transmitted paternally or maternally. Thus, the *CLPG* mutation appeared to inactivate a *cis*-acting silencer element controlling the transcription levels of a core cluster of genes from the *BEGAIN-DIO3* domain in postnatal skeletal muscle, without perturbing their imprinting status (Charlier et al. 2001b; Bidwell et al. 2001, 2004; Smit et al. 2005; Murphy et al. 2005; Perkins et al. 2006; White et al. 2008; Caiment et al. 2010).

That the corresponding dodecamer motif indeed corresponds to a cis-acting regulatory element is further supported by the fact that it maps to a histone-code defined “strong enhancer” signature (Ernst et al. 2011) and that it binds polII and JunD albeit in erythroleukemic K562 cells (Gerstein et al. 2012). However, how this regulatory element operates in skeletal muscle remains largely unknown. When subjecting ~100 bp centered around the mutation to a MatInspector analysis, the first striking observation is that both wild-type and mutant sequence contain E-boxes recognized by myogenic regulatory bHLH-containing factors (MRFs), which might be related to the muscle-specificity of the regulatory element. However, EMSA experiments did not reveal a differential affinity of wild-type and *CLPG* oligonucleotides for MyoD (Freking et al. 2002). Intriguingly, MatInspector analysis suggests that the *CLPG* mutation abrogates a binding site for the Pokemon/ZBTB7A transcriptional repressor, which might be compatible with the observed loss of silencer function.

Whichever the intervening *trans*-acting partners are, the effect of the *CLPG* mutation is accompanied by at least three epigenetic *cis*-alterations: (1) the majority of CpG sites in the immediate vicinity of the *CLPG* mutation are refractory to methylation imposed in postnatal skeletal muscle on the wild-type allele (Murphy et al. 2006; Takeda et al. 2006), (2) the *CLPG* mutation uncovers novel DNase-I hypersensitive sites (DHS) in postnatal skeletal muscle of which one colocalizes nearly exactly with the mutation site (Takeda et al. 2006), and (3) the *CLPG* mutation enhances bidirectional intergenic transcription across the entire *DLK1-GTL2* interval (Takeda et al. 2006).

Hoping to facilitate the study of the *CLPG* phenomenon, we established cultures of myoblasts derived from lambs of the four possible *CLPG* genotypes. However, the *cis*-effect of the mutation on the transcript levels of the neighboring genes, which was so striking *in vivo*, was lost in the cultured myoblasts.

4.6 *Trans*-Effects of the *CLPG* Mutation

Uncovering the *cis*-effect of the *CLPG* mutation revealed at least part of its *modus operandi* yet did not explain polar overdominance. The transcript profile of callipyge animals suggested that ectopic expression of either *DLK1* and/or *RTL1* caused the phenotype, however *CLPG/CLPG* animals shared this feature with their $+^{Mat}/CLPG^{Pat}$ sibs, while being phenotypically wild-type.

The veil was in part lifted when the expression of DLK1 was studied at the protein rather than RNA level. Indeed, despite comparable levels of *DLK1* transcripts in $+^{Mat}/CLPG^{Pat}$ and *CLPG/CLPG* animals, DLK1 protein could only be detected in skeletal muscle of $+^{Mat}/CLPG^{Pat}$ ones. There was thus a perfect correlation between the presence of DLK1 protein in skeletal muscle and phenotype: present in callipyge $+^{Mat}/CLPG^{Pat}$ animals, absent in wild-type $+/+$, *CLPG*^{*Mat*}/*Pat*, and *CLPG/CLPG* animals (Davis et al. 2004). Note that it was

subsequently reported that low levels of DLK1 protein could be detected in skeletal muscle of *CLPG/CLPG* animals (White et al. 2008).

It thus appeared that the *DLK1* and *RTL1* mRNA transcribed from the padumnal *CLPG* allele were posttranscriptionally down-regulated in *CLPG/CLPG* but not in $+^{Mat}/CLPG^{Pat}$ individuals. The noncoding RNAs transcribed from the madumnal *CLPG* allele in *CLPG/CLPG* but not in $+^{Mat}/CLPG^{Pat}$ animals were obviously the best candidate mediators of this Mat-to-Pat *trans* effect. Amongst these, the numerous miRNAs processed from *RTL1-AS* and *MIRG* in particular, stood out as the prime suspects given the known function of miRNA in posttranscriptional inhibition of target mRNAs (Georges et al. 2003).

Because of their perfect sequence complementarity with the intron-less *RTL1* transcripts, the miRNA processed from *RTL1-AS* were excellent candidate *trans*-inhibitors of *RTL1*. Accordingly, we readily identified *RTL1* degradation products terminating at the exact positions predicted for AGO2-mediated “slicing” (i.e., the residue facing nucleotide position 11 of the corresponding miRNA) in skeletal muscle of *CLPG/CLPG* animals (Davis et al. 2005). Thus, in *CLPG/CLPG* animals, miRNAs processed from the *RTL1-AS* pri-miRNA transcribed from the madumnal allele indeed *trans*-inhibit *RTL1* transcripts originating from the padumnal allele, the very mechanism predicted to underlie polar overdominance. The observed RNAi also explained why *RTL1* transcripts, although more abundant in *CLPG/CLPG* and $+^{Mat}/CLPG^{Pat}$ animals when compared to the two other genotypes, were nevertheless less abundant in *CLPG/CLPG* than in $+^{Mat}/CLPG^{Pat}$ animals (Charlier et al. 2001b).

Evidence for the same AGO2-mediated slicing of *RTL1* transcripts mediated by miRNAs derived from *RTL1-AS* was also obtained in placenta of wild-type mice, which is the more likely physiological location of this *trans*-interaction (Davis et al. 2005). It remains one of the very few examples of miRNA-dependent slicing in animals.

The absence of DLK1 protein in *CLPG/CLPG* animals pointed to the occurrence of a similar Mat-to-Pat *trans*-inhibition of *DLK1* transcripts in these animals. Is this *trans*-effect also mediated by miRNAs from the *CLPG* domain? No miRNAs from the *BEGAIN-DIO3* domain appeared to be perfectly complementary to *DLK1* transcripts over their entire length, as observed for *RTL1*. However, it is well established that, in the vast majority of cases, animal miRNAs recognize their targets via fuzzy, partial sequence complementarity that tends to be nucleated by the miRNA “seed” (corresponding to residues 2–8 of the miRNA), and primarily occurs in the 3'UTR. This leads to down-regulation of the target by a combination of splicing-independent mRNA degradation and translational inhibition. The observed ~threefold reduction in *DLK1* transcript levels and >tenfold reduction in DLK1 protein levels in *CLPG/CLPG* animals when compared to $+^{Mat}/CLPG^{Pat}$ animals, was—at first glance—highly reminiscent of the initial report of the mode-of-action of the paradigmatic *lin-4* animal miRNA on its *lin-28* target in *C. elegans* (Seggerson et al. 2002), hence supporting miRNA-dependent down-regulation of *DLK1* in *CLPG/CLPG* animals.

To identify the miRNA from the *BEGAIN-DIO3* domain that might mediated the *trans*-inhibition of *DLK1* in *CLPG/CLPG* animals, we first used next generation sequencing to generate an exhaustive catalogue of miRNAs expressed in *longissimus dorsi* of 8-week-old sheep of the four possible *CLPG* genotypes. We identified a total of 747 miRNA “species.” By species, we refer to a population of “isomirs” (including edited variants) derived from the same arm of a hairpin precursor. Hundred and ten of these species mapped to 74 hairpin loops residing within the *BEGAIN-DIO3* domain (fifty within *MIRG* and five within *RTL1-AS*). We then evaluated whether any one of these miRNAs exhibited an unusual affinity for *DLK1* and might therefore be responsible for the observed Mat-to-Pat *trans*-effect. We first evaluated the affinity of miRNAs for *DLK1* using bioinformatic predictions. None of the miRNAs could be convincingly incriminated using this approach. At best, the whole set of miRNAs from the *BEGAIN-DIO3* domain was showing a marginally significant affinity for the coding region of the *DLK1* transcripts when considered jointly (“as a quadrille”) (Caiment et al. 2010). As sensitivity and specificity of bioinformatics predictions of miRNA-target interactions are reputedly poor, we backed the bioinformatic predictions up with a reporter assay (in cell culture) designed to interrogate the affinity of all miRNA from the domain for the full-length ovine *DLK1* transcripts. Again, not a single miRNAs emerged from this assay as being single-handedly capable of down-regulating *DLK1* to an extent comparable with that observed in vivo (Chen et al. unpublished). We finally performed RNA immunoprecipitation (RIP) experiments using anti-AGO2 antibodies under conditions that allow coimmunoprecipitation of miRNA targets. *DLK1* coimmunoprecipitation was not convincingly increased in skeletal muscle of *CLPG/CLPG* animals when compared to +^{Mat}/*CLPG*^{Pat} animals (Takeda et al. unpublished). Taken together and against expectations, these results did not support a direct role of miRNAs from the *BEGAIN-DIO3* domain in mediating the Mat-to-Pat *trans*-inhibition of *DLK1* observed in *CLPG/CLPG* animals.

These findings obviously raise the question as to what the actual molecular mechanism might be? Are any of the maternally expressed long noncoding RNA genes (such as *GTL2*) directly involved? Is the effect miRNA-dependent but indirect (i.e., the miRNA down-regulate a trans-acting factor that is required for *DLK1* translation)? Further work is obviously required to elucidate this still open question.

In addition to these blatant Mat-to-Pat *trans*-effects on *RTL1* and *DLK1*, detailed examination of transcript and DNA methylation levels revealed two, more subtle *trans*-effects of the *CLPG* mutation. The first corresponds to the consistently observed higher expression levels of all maternally expressed noncoding RNA genes in +^{Mat}/*CLPG*^{Pat} when compared to +/+ and in *CLPG/CLPG* when compared to *CLPG*^{Mat}/+^{Pat} animals. Thus the presence of the *CLPG* mutation on the padumnal chromosome, results—one way or the other—in increased expression levels of the noncoding RNA genes transcribed from the madumnal chromosomes (f.i. Charlier et al. 2001b). The molecular mechanisms underlying this observation remain unknown, but an attractive hypothesis is transvection, i.e., the somewhat

profane suggestion that *cis*-acting regulatory elements may actually sometimes operate in *trans* (f.i. Kennisson and Southworth 2002). This hypothesis can now be tested using Chromosome Conformation Capture, combined with allelic discrimination.

The second subtle *trans*-effect is the observation that, in $+^{Mat}/CLPG^{Pat}$ and $CLPG^{Mat}/+^{Pat}$ heterozygous individuals, the DNA methylation status of the padumnal chromosome (in the immediate vicinity of the *CLPG* mutation), although primarily determined by *CLPG* genotype, is modified by the genotype of the madumnal chromosome. Thus, the padumnal *CLPG* allele is more methylated in $+^{Mat}/CLPG^{Pat}$ than in $CLPG/CLPG$ animals, while the padumnal $+$ allele is less methylated in $CLPG^{Mat}/+^{Pat}$ animals than in $+/+$ animals (Takeda et al. 2006). The significance and mechanisms underlying this observation remain completely unknown.

4.7 From *CLPG* Genotype to Callipyge Phenotype

The previous sections highlight the considerable progress that has been achieved in deciphering the complex effects of the *CLPG* mutation on the expression profiles of a core cluster of flanking genes in animals of the four possible *CLPG* genotypes. But how are these changes in gene expression leading to the callipyge muscular hypertrophy of $+^{Mat}/CLPG^{Pat}$ animals?

One of the most striking characteristics distinguishing $+^{Mat}/CLPG^{Pat}$ animals from the other three genotypes is the ectopic expression of DLK1 protein confined to hypertrophied muscle. To test whether this could be the primary determinant of the hypertrophy, we generated transgenic mice expressing the membrane-bound C form of ovine DLK1 (i.e., the isoform found to be ectopically expressed in callipyge animals) under the dependence of the myosin light chain 3F promoter and 2E enhancer. We obtained two lines with spatiotemporal expression of ovine DLK1 highly reminiscent of that observed on $+^{Mat}/CLPG^{Pat}$ sheep (i.e., ectopic expression in postnatal skeletal muscle). In both lines, adult (11 or 25 weeks old) transgenic animals exhibited a highly significant muscular hypertrophy resulting from a ~10 % increase in myofiber diameter. Taken together, we considered that these results were providing strong support for a direct role of postnatal DLK1 overexpression in mediating the muscular hypertrophy of callipyge sheep (Davis et al. 2004). It is noteworthy that an increase in DLK1 expression respecting the spatiotemporal expression pattern of the endogenous locus (hence prenatal but not postnatal overexpression in skeletal muscle) appears not to affect the development of skeletal muscle (da Rocha et al. 2009), yet that conditional, muscle-specific knockout of *DLK1* causes a reduction in muscle mass (Waddell et al. 2010).

$+^{Mat}/CLPG^{Pat}$ callipyge sheep also differentiate themselves most likely from the three other genotypes by the ectopic expression of *RTL1* (Byrne et al. 2010). Does *RTL1* also contribute to the callipyge phenotype? We have produced transgenic lines expressing ovine *RTL1* under the dependence of the same myosin light chain

3F promoter and 2E enhancer. Preliminary results suggest that ectopic expression of *RTL1* enhances organismal (i.e., multi-organ) growth, including that of muscle. Contrary to callipyge sheep, however, the relative mass of muscle does not seem to be affected. Double transgenic mice (*DLK1* and *RTL1*) are presently being generated to test for a possible synergistic effect between these two genes. It remains possible, therefore, that *RTL1* participates with *DLK1* in causing the callipyge phenotype.

How does ectopic expression of *DLK1* and possible *RTL1* cause the observed muscular hypertrophy, i.e., what are the downstream effectors targeted by *DLK1* and/or *RTL1*? Three studies have used bovine microarrays to compare the transcriptome of affected muscle (f.i. *longissimus dorsi* and *semimembranosus*) between callipyge (+^{Mat}/*CLPG*^{Pat}) and wild-type (+/+) animals (Fleming-Waddell et al. 2007, 2009; Vuocolo et al. 2007). Amongst the genes whose expression was most significantly affected, figured—as expected—a number of myosin heavy chain genes, reflecting the shift towards glycolytic type II fibers in hypertrophied muscle. The expression of the same set of genes was predicted to be altered when comparing semimembranosus and longissimus dorsi (predominantly fast oxidative) with semitendinosus (predominantly fast glycolytic) in wild-type (+/+) animals, and this was used in one of the studies to identify a set of differentially expressed genes specific for the SD (+^{Mat}/*CLPG*^{Pat}) versus SD (+/+) contrast. Across the three studies, more than 300 genes were identified whose expression might be altered in hypertrophied muscles of callipyge animals. Slightly more than halve of these would be overexpressed in callipyge muscle. Analysis of the gene lists didn't clearly reveal the predominant role of specific pathways, yet the authors highlighted a possible involvement of histone modifying enzymes, and of the AKT/mTOR signaling pathway. It is noteworthy that the skeletal muscle transcriptome of *CLPG*^{Mat}/+^{Pat} animals (which are ectopically expressing the maternally expressed noncoding RNAs) was virtually not altered when compared to that of +/+ animals. The secondary events leading to the callipyge muscular hypertrophy therefore remain largely unknown.

4.8 Polar Overdominance: Sheep Idiosyncrasy or Common Phenomenon?

There is at least one other mammalian phenotype that is inherited in a manner reminiscent of polar overdominance: the DDK syndrome (recognized as early as 1967). The main feature of the DDK syndrome is the nearly fully penetrant early lethality of embryos derived from crosses between females from the DDK strain with males from most non-DDK strains, while the reciprocal cross and the respective within strain matings are fully fertile. It is now known that the embryonic lethality is due to the incompatibility between a cytoplasmic factor contributed by the DDK ooplasm (possibly an RNA) and a nuclear factor contributed by the

non-DDK sperm cell. The two factors are encoded by a pair of closely linked genes mapping to the *Om* (*Ovum mutant*) locus on mouse chromosome 11. The parent-of-origin effect characterizing the DDK syndrome therefore results from the involvement of genes expressed in either the male or female germ line rather than from the involvement of imprinted genes (reviewed in Wakasugi (2007)).

Recently, Wolf et al. (2008) scanned the mouse genome for quantitative trait loci (QTL) influencing weight and growth in a purpose-build three-generation design allowing distinction of alternative heterozygotes and hence testing for parent-of-origin effects independent of maternal effects. They detected multiple supposedly imprinted QTL (iQTL), including several characterized by polar overdominance reminiscent of the callipyge phenomenon, as well as iQTL with polar underdominance as well as bipolar dominance (the two heterozygous genotypes differ from each other, while the homozygous genotypes do not). These results suggest that polar overdominance and related inheritance patterns might be more common than generally appreciated.

Kong et al. (2009) revisited seven risk loci identified by GWAS for common complex diseases in human, and mapping in the vicinity of known imprinted gene clusters, accounting for the parental origin of an individual's haplotypes, which allowed them to test for imprinting effects. Significant parent-of-origin effects were observed for five of the seven examined loci. With exception of one of these loci, it is somewhat unclear whether the authors specifically tested for polar overdominance-like effects, and it therefore remains difficult to conclude from this study in how far such modes of inheritance might contribute to inherited risk for common complex diseases.

One study examined association between common variants in the *DLK1-GTL2* region and childhood obesity using trio data allowing distinction between the two classes of heterozygotes, and therefore testing of parent-of-origin effects including polar overdominance (Wermter et al. 2008). Intriguingly, modeling parent-of-origin effects revealed an association that was overlooked when using a standard Mendelian model, and polar overdominance appeared to better explain the data than a simple imprinting model. However, the effect was strongest for rs1802710, a synonymous variant in the *DLK1* ORF, which remains difficult to reconcile with the complex regulatory effects observed for the ovine *CLPG* mutation. Follow-up studies are needed to confirm and extend these intriguing observations.

There have been at least two reports claiming the detection in line-crosses of QTL affecting either growth, fatness, or muscle mass, with polar overdominance-like effects in the vicinity of the porcine *BEGAIN-DIO3* region (Kim et al. 2004; Boysen et al. 2010). It is important to note, however, that the experimental designs and analyses methods used in both studies are now known to be prone to detection of false-positive imprinting effects. This is primarily due to the fact that the analyses methods assume that alternative QTL alleles (Q and q) are fixed in the parental lines, and that all F1 parents are therefore assumed to be of Qq QTL genotype. It was shown that such models are very often not accurate and cause so-called pseudo-imprinting (f.i. Sandor and Georges 2008). Hence, further work is needed to evaluate the widespread importance of polar overdominance in livestock.

4.9 Conclusion

Analysis of the callipyge phenotype and its unusual polar overdominant inheritance has revealed unique molecular mechanism including, in particular, the miRNA-mediated interaction between the maternal and paternal homologues at the imprinted *BEGAIN-DIO3* domain. Much remains to be learned, however, about the ultimate causes of polar overdominance: what are the *trans*-acting factors mediating the *cis*-effect of the *CLPG* mutation and how do they exert their effect? What are the mediators of the Mat-to-Pat *trans*-inhibition of *DLK1*, the most likely effector of the callipyge muscular hypertrophy? What are the secondary mechanisms by which ectopic expression of the callipyge effectors exert their effect? What explains the rostro-caudal phenotypic gradient? Moreover, a more detailed understanding of the callipyge phenomenology may contribute to a better understanding of the *BEGAIN-DIO3* domain and in particular of the role of the lincRNAs expressed from the madumal chromosome: what are the roles of *GTL2*, *RTL1-AS*, *RIAN*, *MIRG*, and the C/D snoRNAs and miRNAs derived from them? What are the different *cis*-elements regulating the expression of the genes in the *BEGAIN-DIO3* domain and how do they interact? Where does the Mat-to-Pat *trans*-interaction, revealed by studying the callipyge phenotype, occur in wild-type animals and what is its biological significance?

Acknowledgments Callipyge work at the Unit of Animal Genomics was funded by grants from the Belgian FNRS, the Communauté Française de Belgique (ARC Mirage), and the European Union (FP7 Callimir project).

References

- Berghmans S, Segers K, Shay T, Georges M, Cockett N, Charlier C (2001) Breakpoint mapping positions the callipyge gene within a 450-kilobase chromosome segment containing the *DLK1* and *GTL2* genes. *Mamm Genome* 12:183–185
- Bidwell CA, Shay TL, Georges M, Beever JE, Berghmans S, Cockett NE (2001) Differential expression of the *GTL2* gene within the callipyge region of ovine chromosome 18. *Anim Genet* 32:248–256
- Bidwell CA, Kramer LN, Perkins AC, Hadfield TS, Moody DE, Cockett NE (2004) Expression of *PEG11* and *PEG11AS* transcripts in normal and callipyge sheep. *BMC Biol* 2:17
- Boysen TJ, Tetens J, Thaller G (2010) Detection of a quantitative trait locus for ham weight with polar overdominance near the ortholog of the callipyge locus in an experimental pig F2 population. *J Anim Sci* 88:3167–3172
- Bray SJ, Takada S, Harrison E, Shen SC, Ferguson-Smith AC (2008) The atypical mammalian ligand Delta-like homologue 1 (*Dlk1*) can regulate Notch signalling in *Drosophila*. *BMC Dev Biol* 8:11
- Byrne K, Colgrave ML, Vuocolo T, Pearson R, Bidwell CA, Cockett NE, Lynn DJ, Fleming-Waddell JN, Tellam RL (2010) The imprinted retrotransposon-like gene *PEG11* (*RTL1*) is expressed as a full-length protein in skeletal muscle from Callipyge sheep. *PLoS One* 5:e8638

- Caiment F, Charlier C, Hadfield T, Cockett N, Georges M, Baurain D (2010) Assessing the effect of the CLPG mutation on the microRNA catalog of skeletal muscle using high-throughput sequencing. *Genome Res* 20:1651–1662
- Carpenter CE, Rice OD, Cockett NE, Snowden GD (1996) Histology and composition of muscles from normal and callipyge lambs. *J Anim Sci* 74:388–393
- Cavaillé J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachelier JP (2002) Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet* 11:1527–1538
- Charalambous M, Ferron SR, da Rocha ST, Murray AJ, Rowland T, Ito M, Schuster-Gossler K, Hernandez A, Ferguson-Smith AC (2012) Imprinted gene dosage is critical for the transition to independent life. *Cell Metab* 15:209–221
- Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G, Georges M (2001a) Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (*clpg*) locus and identification of six imprinted transcripts: *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*. *Genome Res* 11:850–862
- Charlier C, Segers K, Karim L, Shay T, Gyapay G, Cockett N, Georges M (2001b) The callipyge mutation enhances the expression of coregulated imprinted genes in *cis* without affecting their imprinting status. *Nat Genet* 27:367–369
- Cockett NE, Jackson SP, Shay TL, Nielsen D, Moore SS, Steele MR, Barendse W, Green RD, Georges M (1994) Chromosomal localization of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc Natl Acad Sci U S A* 91:3019–3023
- Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowden GD, Nielsen DM, Georges M (1996) Polar overdominance at the ovine callipyge locus. *Science* 273:236–238
- da Rocha ST, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC (2008) Genomic imprinting at the mammalian *Dlk1-Dio3* domain. *Trends Genet* 24:306–316
- da Rocha ST, Charalambous M, Lin SP, Gutteridge I, Ito Y, Gray D, Dean W, Ferguson-Smith AC (2009) Gene dosage effects of the imprinted delta-like homologue 1 (*dlk1/pref1*) in development: implications for the evolution of imprinting. *PLoS Genet* 5:e1000392
- Davis E, Jensen CH, Schroder HD, Farnir F, Shay-Hadfield T, Kliem A, Cockett N, Georges M, Charlier C (2004) Ectopic expression of *DLK1* protein in skeletal muscle of padminal heterozygotes causes the callipyge phenotype. *Curr Biol* 14:1858–1862
- Davis E, Caiment F, Tordoir X, Cavaillé J, Ferguson-Smith A, Cockett N, Georges M, Charlier C (2005) RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus. *Curr Biol* 15:743–749
- Deguchi M, Hata Y, Takeuchi M, Ide N, Hirao K, Yao I, Irie M, Toyoda A, Takai Y (1998) BEGAIN (brain-enriched guanylate kinase-associated protein), a novel neuronal PSD-95/SAP90-binding protein. *J Biol Chem* 273:26269–26272
- Edwards CA, Mungall AJ, Matthews L, Ryder E, Gray DJ, Pask AJ, Shaw G, Graves JA, Rogers J, SAVOIR consortium, Dunham I, Renfree MB, Ferguson-Smith AC (2008) The evolution of the *DLK1-DIO3* imprinted domain in mammals. *PLoS Biol* 6:e135
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49
- Fahrenkrug SC, Freking BA, Rexroad CE III, Leymaster KA, Kappes SM, Smith TP (2000) Comparative mapping of the ovine *clpg* locus. *Mamm Genome* 11:871–876
- Falix FA, Aronson DC, Lamers WH, Gaemers IC (2012) Possible roles of *DLK1* in the Notch pathway during development and disease. *Biochim Biophys Acta* 1822:988–995
- Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* 12:565–575
- Ferrón SR, Charalambous M, Radford E, McEwen K, Wildner H, Hind E, Morante-Redolat JM, Laborda J, Guillemot F, Bauer SR, Fariñas I, Ferguson-Smith AC (2011) Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* 475:381–385

- Fleming-Waddell JN, Wilson LM, Olbricht GR, Vuocolo T, Byrne K, Craig BA, Tellam RL, Cockett NE, Bidwell CA (2007) Analysis of gene expression during the onset of muscle hypertrophy in callipyge lambs. *Anim Gene* 38:28–36
- Fleming-Waddell JN, Olbricht GR, Taxis TM, White JD, Vuocolo T, Craig BA, Tellam RL, Neary MK, Cockett NE, Bidwell CA (2009) Effect of DLK1 and RTL1 but not MEG3 or MEG8 on muscle gene expression in Callipyge lambs. *PLoS One* 4:e7399
- Freking BA, Keele JW, Beattie CW, Kappes SM, Smith TP, Sonstegard TS, Nielsen MK, Leymaster KA (1998a) Evaluation of the ovine callipyge locus: I. Relative chromosomal position and gene action. *J Anim Sci* 76:2062–2071
- Freking BA, Keele JW, Nielsen MK, Leymaster KA (1998b) Evaluation of the ovine callipyge locus: II. Genotypic effects on growth, slaughter, and carcass traits. *J Anim Sci* 76:2549–2559
- Freking BA, Keele JW, Shackelford SD, Wheeler TL, Koohmaraie M, Nielsen MK, Leymaster KA (1999) Evaluation of the ovine callipyge locus: III. Genotypic effects on meat quality traits. *J Anim Sci* 77:2336–2344
- Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL, Smith TP (2002) Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res* 12:1496–1506
- Freking BA, Leymaster KA (2006) Evaluation of the ovine callipyge locus: IV. Genotypic effects on reproductive traits. *J Anim Sci* 84:311–316
- Georges M, Charlier C, Cockett N (2003) The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet* 19:248–252
- Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Harmanai A, Jain P, Kasowski M, Lacroute P, Leng J, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature* 489:91–100
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202
- Girardot M, Cavaillé J, Feil R (2012) Small regulatory RNAs controlled by genomic imprinting and their contribution to human disease. *Epigenetics* 7:1341–1348
- Glazov EA, McWilliam S, Barris WC, Dalrymple BP (2008) Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. *Mol Biol Evol* 25:939–948
- Haig D (2000) The kinship theory of genomic imprinting. *Annu Rev Ecol Syst* 31:9–32
- Jackson SP, Green RD, Miller MF (1997a) Phenotypic characterization of ramboillet sheep expressing the callipyge gene: I. Inheritance of the condition and production characteristics. *J Anim Sci* 75:14–18
- Jackson SP, Miller MF, Green RD (1997b) Phenotypic characterization of Ramboillet sheep expressing the callipyge gene: II. Carcass characteristics and retail yield. *J Anim Sci* 75:125–132
- Jackson SP, Miller MF, Green RD (1997c) Phenotypic characterization of ramboillet sheep expression the callipyge gene: III. Muscle weights and muscle weight distribution. *J Anim Sci* 75:133–138
- Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315:1137–1140
- Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T,

- Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, DIAGRAM Consortium, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K (2009) Parental origin of sequence variants associated with complex diseases. *Nature* 462:868–874
- Kennison JA, Southworth JW (2002) Transvection in *Drosophila*. In: Dunlap JC, Wu CT (ed) Homology effects. Academic Press, pp 399–420
- Kim KS, Kim JJ, Dekkers JC, Rothschild MF (2004) Polar overdominant inheritance of a DLK1 polymorphism is associated with growth and fatness in pigs. *Mamm Genome* 15:552–559
- Koohmaraie M, Shackelford SD, Wheeler TL, Lonergan SM, Doumit ME (1995) A muscle hypertrophy condition in lamb (callipyge): characterization of effects on muscle growth and meat quality traits. *J Anim Sci* 73:3596–3607
- Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaillé J, Ferguson-Smith AC (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* 35:97–102
- Lin SP, Coan P, da Rocha ST, Seitz H, Cavaillé J, Teng PW, Takada S, Ferguson-Smith AC (2007) Differential regulation of imprinting in the murine embryo and placenta by the Dlk1-Dio3 imprinting control region. *Development* 134:417–426
- Mirshekar-Syahkal B, Haak E, Kimber GM, van Leusden K, Harvey K, O'Rourke J, Laborda J, Bauer SR, de Bruijn MF, Ferguson-Smith AC, Dzierzak E, Ottersbach K (2013) Dlk1 is a negative regulator of emerging hematopoietic stem and progenitor cells. *Haematologica* 98:163–171
- Murphy SK, Freking BA, Smith TP, Leymaster K, Nolan CM, Wylie AA, Evans HK, Jirtle RL (2005) Abnormal postnatal maintenance of elevated DLK1 transcript levels in callipyge sheep. *Mamm Genome* 16:171–183
- Murphy SK, Nolan CM, Huang Z, Kucera KS, Freking BA, Smith TP, Leymaster KA, Weidman JR, Jirtle RL (2006) Callipyge mutation affects gene expression in cis: a potential role for chromatin structure. *Genome Res* 16:340–346
- Paulsen M, Takada S, Youngson NA, Benchaib M, Charlier C, Segers K, Georges M, Ferguson-Smith AC (2001) Comparative sequence analysis of the imprinted Dlk1-Gtl2 locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the Igf2-H19 region. *Genome Res* 11:2085–2094
- Perkins AC, Kramer LN, Spurlock DM, Hadfield TS, Cockett NE, Bidwell CA (2006) Postnatal changes in the expression of genes located in the callipyge region in sheep skeletal muscle. *Anim Genet* 37:535–542
- Sandor C, Georges M (2008) On the detection of imprinted quantitative trait loci in line crosses: effect of linkage disequilibrium. *Genetics* 180:1167–1175
- Schuster-Gossler K, Bilinski P, Sado T, Ferguson-Smith A, Gossler A (1998) The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Dev Dyn* 212:214–228
- Seggerson K, Tang L, Moss EG (2002) Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev Biol* 243:215–225
- Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachelier JP, Ferguson-Smith AC, Cavaillé J (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet* 34:261–262
- Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaillé J (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res* 14:1741–1748
- Sehara Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, Hino T, Suzuki-Migishima R, Kohda T, Ogura A, Ogata T, Yokoyama M, Kaneko-Ishino T, Ishino F (2008) Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta. *Nat Genet* 40:243–248
- Shay TL, Berghmans S, Segers K, Meyers S, Beever JE, Womack JE, Georges M, Charlier C, Cockett NE (2001) Fine-mapping and construction of a bovine contig spanning the ovine callipyge locus. *Mamm Genome* 12:141–149

- Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, Gyapay G, Snowder G, Georges M, Cockett N, Charlier C (2003) Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* 163:453–456
- Smit MA, Tordoir X, Gyapay G, Cockett NE, Georges M, Charlier C (2005) BEGAIN: a novel imprinted gene that generates paternally expressed transcripts in a tissue- and promoter-specific manner in sheep. *Mamm Genome* 16:801–814
- Takada S, Paulsen M, Tevendale M, Tsai CE, Kelsey G, Cattanach BM, Ferguson-Smith AC (2002) Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum Mol Gene* 11:77–86
- Takeda H, Caiment F, Smit M, Hiard S, Tordoir X, Cockett N, Georges M, Charlier C (2006) The callipyge mutation enhances bidirectional long-range DLK1-GTL2 intergenic transcription in cis. *Proc Natl Acad Sci U S A* 103:8119–8124
- Tierling S, Gasparoni G, Youngson N, Paulsen M (2009) The Begain gene marks the centromeric boundary of the imprinted region on mouse chromosome 12. *Mamm Genome* 20:699–710
- Tsai CE, Lin SP, Ito M, Takagi N, Takada S, Ferguson-Smith AC (2002) Genomic imprinting contributes to thyroid hormone metabolism in the mouse embryo. *Curr Biol* 12:1221–1226
- Vuocolo T, Byrne K, White J, McWilliam S, Reverter A, Cockett NE, Tellam RL (2007) Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. *Physiol Genomics* 28:253–272
- Waddell JN, Zhang P, Wen Y, Gupta SK, Yevtodiyeenko A, Schmidt JV, Bidwell CA, Kumar A, Kuang S (2010) Dlk1 is necessary for proper skeletal muscle development and regeneration. *PLoS One* 5:e15055
- Wakasugi N (2007) Embryologic, cytobiologic and genetic interpretations of DDK syndrome in mice. *Dev Growth Differ* 49:55–559
- Wermter AK, Scherag A, Meyre D, Reichwald K, Durand E, Nguyen TT, Koberwitz K, Lichtner P, Meitinger T, Schäfer H, Hinney A, Froguel P, Hebebrand J, Brönnert G (2008) Preferential reciprocal transfer of paternal/maternal DLK1 alleles to obese children: first evidence of polar overdominance in humans. *Eur J Hum Genet* 16:1126–1134
- White JD, Vuocolo T, McDonagh M, Grounds MD, Harper GS, Cockett NE, Tellam R (2008) Analysis of the callipyge phenotype through skeletal muscle development; association of Dlk1 with muscle precursor cells. *Differentiation* 76:283–298
- Wolf JB, Cheverud JM, Roseman C, Hager R (2008) Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet* 4:e1000091
- Youngson NA, Kocialkowski S, Peel N, Ferguson-Smith AC (2005) A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. *J Mol Evol* 61:481–490
- Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M, Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* 40:939–953

Chapter 5

Transgenerational Epigenetic Effects and Complex Inheritance Patterns

Anna K. Naumova

Abstract Mendel's laws describing the inheritance of simple genetic traits are based on the assumption that genes are transmitted unchanged from parents to offspring. However, this assumption is not always correct as the mammalian genome undergoes massive epigenetic reprogramming during gametogenesis and early embryonic development. Furthermore, stochastic or environmentally induced epigenetic variation may lead to situations where the epigenetic marking of the same allele differs between parents and offspring, among siblings or even monozygotic twins. Epigenetic marks may modify the penetrance of a phenotype and cause "non-Mendelian" inheritance even when the causal genetic variant is transmitted from parent to offspring in perfect Mendelian proportions. In this chapter we focus on a particular type of inheritance where epigenetic marks that are present in parental somatic cells fail to be reset in a proportion of parental germ cells and are transmitted to offspring. Such a transgenerational "epigenetic memory" increases the complexity of the inheritance pattern and therefore is of particular interest for geneticists.

A.K. Naumova, Ph.D. (✉)

Departments of Obstetrics and Gynecology and Human Genetics,
McGill University, Montreal, QC, Canada

The Research Institute of the McGill University Health Centre,
McGill University, Montreal, QC, Canada
e-mail: anna.naumova@mcgill.ca

5.1 Epigenetic Remodeling of the Genome in Germ Cells and Early Embryos

Current understanding of the multistep epigenetic remodeling of the genome that takes place in the mammalian germ line is based largely on genome-wide analyses of DNA methylation levels and detailed studies of imprinted genes (for detailed description of epigenetic reprogramming in the mammalian germ line see Chap. 1).

Primordial germ cells are the earliest precursors of an individual's gametes. They originate from the somatic cells of primary ectoderm of the embryo and migrate to the future gonad (genital ridge). In mice, they populate the genital ridge between days 11 and 13 (midgestation) of embryonic development (Ginsburg et al. 1990; Gomperts et al. 1994). In humans, they appear at the genital ridge between 4 and 6 weeks post conception, i.e., during the first trimester of pregnancy (Clark 2007). After migration of primordial germ cells to the genital ridge, a wave of genome-wide DNA demethylation erases the somatic cell-specific DNA methylation marks from their genomes (Szabo and Mann 1995; Hajkova et al. 2002; Feng et al. 2010) (Fig. 5.1). This prepares the primordial germ cell genome for the acquisition of gamete-specific epigenetic marks, which depend on the sex of the embryo. In mammals, male and female gametogenesis differ in many aspects including the developmental timing, length of certain stages, gene expression profiles, RNA storage, and protein accumulation as well as the stages when the genomic DNA becomes methylated. In males, a massive gain of genomic DNA methylation occurs before or at the spermatogonial stage, preceding the onset of meiosis (Davis et al. 1999, 2000; Kaneda et al. 2004; Trasler 2009). Spermatogonia, as well as cells at later stages of spermatogenesis, are constantly renewed and produced during a male's life. In females, the onset of prophase I of meiosis takes place before birth. Oocytes reach the diplotene stage of meiotic prophase and are stored at this stage for months (in mice) or decades (in humans and other mammals with a longer reproductive life span). After puberty, every estrus cycle several oocytes are recruited into the growing oocyte pool. They grow and complete the first meiotic division, whereas the second meiotic division occurs only after fertilization (reviewed in Amleh et al. 2012). During oocyte growth and maturation genome-wide acquisition of DNA methylation occurs (Obata and Kono 2002; Lucifero et al. 2004; Hiura et al. 2006; Smallwood et al. 2011).

The sex-specific epigenetic marks that are established during gametogenesis result in genome-wide epigenetic differences between the mature spermatozoan and oocyte genomes and consequently genome-wide differences between the maternally and paternally derived chromosomes in the embryo (Fig. 5.1) (for details see Chap. 1). After fertilization, the first steps of embryonic development are marked by genome-wide loss of DNA methylation (Howlett and Reik 1991; Monk 1995) and chromatin reorganization (reviewed in Feng et al. 2010). The consequence of this wave of DNA demethylation is loss of parent-of-origin-dependent DNA methylation

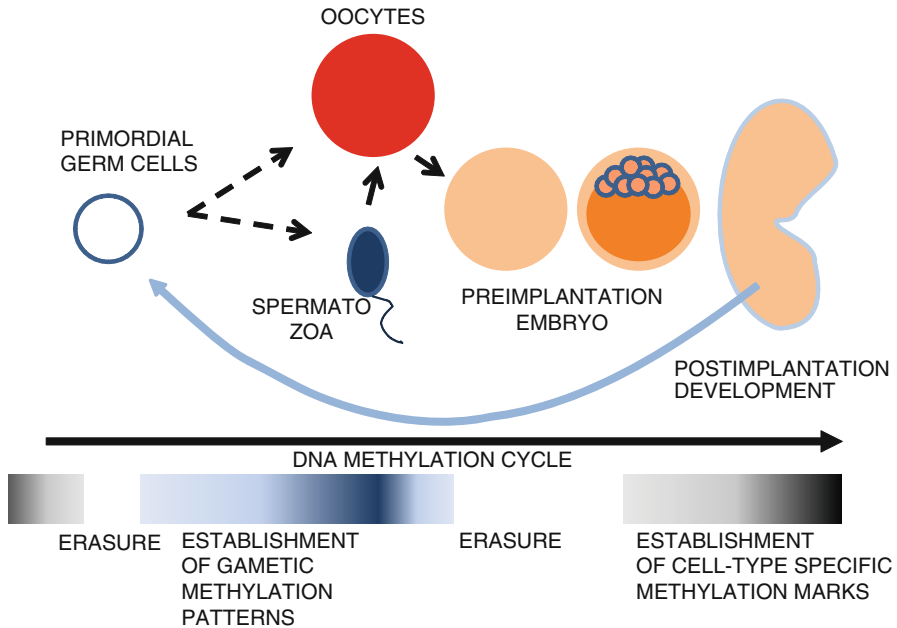


Fig. 5.1 The cycle of epigenetic reprogramming. DNA methylation and demethylation in the mammalian germ line and embryo

at most genomic loci. After global erasure of the vast majority of gametic DNA methylation marks, a new wave of DNA methylation occurs during and after the blastocyst stage (Monk 1995). From this point on, cell type-specific DNA methylation patterns that are essential for the maintenance of tissue-specific gene expression are established (Feng et al. 2010).

Thus in mammals, there are two waves of genome-wide DNA demethylation that occur during the time window which is particularly important for genetic transmission, in primordial germ cells and in early preimplantation embryos; and at least one wave of DNA methylation that occurs during gametogenesis. Based on the cycle of DNA demethylation and remethylation, one may predict that anomalies in methylation reprogramming could cause dramatic changes in the epigenetic marking of the gametic or embryonic genomes, and thereby influence inheritance of genetic traits. It follows therefore that genetic or environmental factors that cause subtle shifts in the global or locus-specific DNA methylation levels specifically in the germ cells of the developing embryo will not affect the individual's somatic cells: their effects will come to light only in the next generation.

5.2 Transgenerational Epigenetic Inheritance. Transgenes and Retroviral Elements

The definition of transgenerational epigenetic inheritance is often a matter of debate (Chong et al. 2007a, b) and several excellent reviews have been recently published on the subject (Daxinger and Whitelaw 2012; Grossniklaus et al. 2013; Lim and Brunet 2013). In this chapter, the term “transgenerational epigenetic inheritance” refers to inheritance of the epigenetic marks of the somatic cells of the parent by his/her offspring, through the germ line (Morgan et al. 1999; Chong et al. 2007a). Emerging evidence implicates specific classes of microRNAs in transgenerational epigenetic effects (Nelson and Nadeau 2010; Ashe et al. 2012; Daxinger and Whitelaw 2012; Grentzinger et al. 2012). However, in this chapter we focus on the currently best understood epigenetic mark—DNA methylation. For transgenerational epigenetic inheritance to occur, the somatic DNA methylation marks must resist the two waves of demethylation, in the primordial germ cells and in preimplantation embryos, as well as the waves of remethylation during gametogenesis and embryonic development (Fig. 5.1). In plants, this is a common occurrence (Mathieu et al. 2007; Feng et al. 2010). In mammals, very few *bona fide* cases of transgenerational epigenetic inheritance have been documented in detail. Nevertheless the phenomenon may be of greater importance than previously appreciated as several recent genome-wide studies of DNA methylation in mouse gametes and embryos demonstrated that certain classes of DNA sequences including endogenous retroviruses resist genome-wide demethylation in primordial germ cells (Hackett et al. 2013) and that methylation levels of CpG islands in preimplantation embryos depend upon their methylation levels in oocytes (Smallwood et al. 2011; Guibert et al. 2012). These recent data demonstrate that epigenetic reprogramming is not equally efficient at all genomic regions/germ cells, providing direct evidence for the hypothesis that was first proposed a couple of decades ago based on non-Mendelian inheritance patterns at several genomic loci (Laird 1987; Follette and Laird 1992; Naumova and Sapienza 1994; Naumova et al. 2001).

Apart from its role in cell differentiation and the orchestration of genome-wide gene regulation, DNA methylation acts as a host-response mechanism protecting the integrity of cellular genomes from foreign DNA invasion (reviewed in Bestor 1998; Matzke and Matzke 1998). Mammalian genomes harbor thousands of retroviral transposable elements, also termed endogenous retroviruses. Their expression is critical for transposition and spreading throughout the genome (reviewed in Hancks and Kazazian 2012), whereas methylation of endogenous retroviruses prevents their expression and thereby maintains genome integrity (Bourc’his and Bestor 2004; Maksakova et al. 2008). This host defense mechanism also targets transgenic sequences that are inserted into the mouse genome for research purposes, thereby often hampering transgenic research (Reik et al. 1987; Sapienza et al. 1987; McGowan et al. 1989). The phenomenon of transgene methylation was investigated in detail in several transgenic mouse lines (Reik et al. 1987; Swain et al. 1987; McGowan et al. 1989; Allen et al. 1990; Pickard et al. 2001; Valenza-Schaerly et al. 2001).

These studies demonstrated that transgene methylation may depend upon the parent of origin (Reik et al. 1987; McGowan et al. 1989) or the mouse strain (Weng et al. 1995; Engler et al. 1998). Moreover, in certain transgenic lines the transgenic DNA sequence may accumulate DNA methylation over generations (Allen et al. 1990), becoming increasingly methylated with every germ line transmission. This indicates that the transgenic sequence resisted the waves of demethylation in primordial germ cells and/or early embryos (Kearns et al. 2000), but attracted methylation during gametogenesis and perhaps at later stages of embryonic development. A series of in-depth studies of methylation of retroviral elements revealed a similar epigenetic response to endogenous retroviruses (reviewed in Rakyan et al. 2002).

The intracisternal A particle (IAP) is a mobile DNA element that contains a full retroviral genome flanked by two long terminal repeats (LTR) (Kuff and Lueders 1988). Insertion of an IAP near or within a gene may cause changes in gene regulation because the IAP harbors promoters that may overtake regulation of expression of neighboring genes. Methylation of the IAP promoter prevents its adverse effect on neighboring genes. Two animal models on which most of our current understanding of transgenerational epigenetic inheritance is based are (1) the insertion of an IAP upstream of the agouti locus (Perry et al. 1994) giving rise to the agouti viable yellow allele (A^{vy}) (Dickies 1962) and (2) the insertion of an IAP into intron 6 of the *Axin 1* gene (Vasicek et al. 1997) giving rise to the *Axin fused* allele ($Axin^{Fu}$) (Reed 1937). In the case of the dominant A^{vy} allele, the IAP LTR drives the expression of the agouti transcript bypassing tissue-specific regulatory elements and causing expression of the agouti gene in cells and at stages where normally the gene is not expressed (Morgan et al. 1999). The agouti gene product is involved in melanocortin receptor binding. Loss of proper cell-specific regulation of the agouti locus has a pleiotropic effect on multiple tissues (Wolff 1965) including the hair follicles where it results in an easily scored phenotype—a yellow coat color. In heterozygous A^{vy} mice DNA methylation levels of the IAP LTR vary between individual mice and even between individual cells (Morgan et al. 1999). Mice with non-methylated IAP have a yellow coat and develop obesity. Hypermethylation of the IAP restores normal regulation of the agouti gene and produces a pseudo-wild-type phenotype with brownish colored fur (termed pseudoagouti) (Morgan et al. 1999). Variation in IAP methylation among individual cells results in mosaic expression of the agouti gene and therefore in a mosaic, also termed mottled, coat color (Fig. 5.2). Thus, a single mutation causes a range of coat color phenotypes. In every generation, about 20 % of the offspring of heterozygous male carriers of the A^{vy} allele are pseudoagouti, independent of the color of the father, i.e., independent of the methylation level of the allele in the parental somatic cells (Wolff 1978; Morgan et al. 1999) (Fig. 5.2a). However, the IAP methylation level in the mother and hence maternal coat color provide a good prediction for the distribution of the coat color among her offspring: pseudoagouti mothers produce more pseudoagouti pups compared to yellow mothers (Morgan et al. 1999) (Fig. 5.2b–d). Moreover, pups born to a pedigree with a pseudoagouti mother and maternal grandmother, have the highest chance of wearing pseudoagouti coats (Morgan et al. 1999) (Fig. 5.2d). Such a correlation between the methylation levels

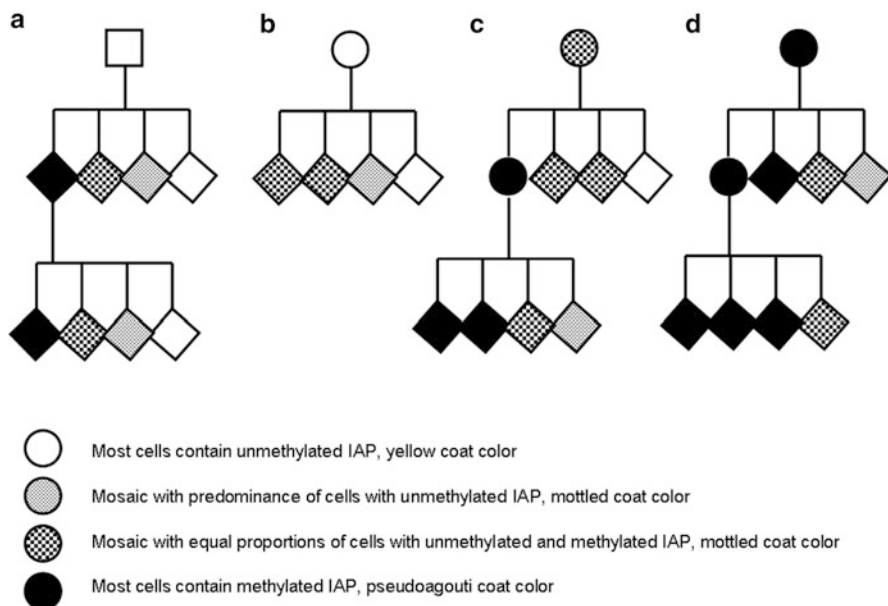


Fig. 5.2 Stochastic epigenetic variation or transgenerational epigenetic inheritance. The diagram depicts the inheritance of the coat color phenotypes in mouse carriers of the agouti viable A^{vy} allele. (a) Stochastic variation in methylation in the offspring does not depend upon the allelic methylation of the IAP LTR in the father. Paternal methylation marks are fully erased after fertilization, but their establishment in the embryo is variable and subject to stochastic influences. (b–d) Transgenerational epigenetic inheritance of allelic DNA methylation and resulting phenotypes. The individual's coat color depends on methylation of the IAP LTR in the mother. DNA methylation is not fully erased in the germ line of the mother and influences the LTR DNA methylation level in the cells of her offspring; therefore mothers with higher methylation levels have a higher chance of producing offspring with high methylation levels and **vice versa**

of the IAP in the somatic cells of the mother and the distribution of methylation levels among her offspring results from failure to reset the somatic DNA methylation marks at the IAP in the A^{vy} allele during gametogenesis and preimplantation development (Rakyan et al. 2003; Blewitt et al. 2006). Thus, the reprogramming defect adds another layer to the inheritance of the genotype—inheritance of the epigenotype or transgenerational epigenetic inheritance. It is important to note that in this model, the fact that inbred mice have virtually identical genomes excludes the possibility that the trait is oligogenic.

Similar to the A^{vy} allele, the $Axin^{Fu}$ mutation results from an insertion of the IAP into intron 6 of the *Axin 1* gene located on mouse chromosome 17 (Vasicek et al. 1997). The LTR of the IAP acts as a promoter and directs the synthesis of several abnormal transcripts. Abnormal transcription products interfere with the normal function of the AXIN1 protein and cause vertebral fusion and a kinky tail. Consequently, in the homozygous state, the mutation causes multiple anomalies and is embryonic lethal (Reed 1937; Dunn and Caspari 1942). However, in the heterozygous state it is dominant with variable penetrance, i.e., a proportion of

heterozygous carriers display mutant phenotypes (kinky tails) whereas other carriers have normal tails (Reed 1937). Initially, the mode of inheritance was defined as semidominant (Reed 1937); however a better understanding of the mechanism supports a dominant mode of inheritance with reduced penetrance. Similar to the A^{vy} mutation, loss of penetrance of this dominant mutation is due to methylation of the LTR promoter which prevents abnormal transcription in the locus (Rakyan et al. 2003). The inheritance pattern of the kinked tail phenotype is consistent with the transgenerational epigenetic inheritance paradigm (Rakyan et al. 2003). An in-depth analysis of the methylation changes during the developmental window between fertilization and blastocyst stage that is critical for epigenetic reprogramming demonstrated that methylation of the IAP is lost during preimplantation development, but reestablished after implantation in accordance with the methylation of the gametes that gave rise to the embryo. Hence, another epigenetic mark, distinct from DNA methylation, directs the reestablishment of DNA methylation patterns at the IAP elements. Several candidates, including a specific profile of histone modifications and presence of hydroxymethylcytosine, are proposed as guides for the remethylation of the IAP after implantation (Fernandez-Gonzalez et al. 2010; Hackett et al. 2013). Thus, the sum of evidence indicates that variation in the IAP methylation patterns and thereby the penetrance of A^{vy} and $Axin^{Fu}$ mutations stem from two sources. Firstly, the somatic epigenetic marks of the paternal IAP LTR alleles are not erased in the primordial germ cells, nor do the alleles show substantial gain of methylation during gametogenesis (Lane et al. 2003; Blewitt et al. 2006). Therefore the embryo inherits almost exactly the same methylation pattern that was in the oocyte/spermatocyte of its parent. Secondly, stochastic gain or loss of methylation occurs while the IAP LTR is being remethylated in the embryo or in the course of multiple cell divisions. This results in the embryo being mosaic for the epigenotype as well as for the phenotype.

Thus, a single mutation (IAP insertion, as demonstrated in the cases of the A^{vy} and $Axin^{Fu}$ alleles) in combination with variable DNA methylation is capable of producing a range of phenotypes. Such an interaction between genetic and epigenetic factors generates a quantitative genetic trait or a dominant trait with variable penetrance. Moreover, failure to reset the IAP methylation marks in the germ line adds another layer of inheritance—the epigenetic inheritance where the somatic methylation marks are transmitted through generations. The mutant allele predisposes to a specific phenotype (coat color, obesity, or kinky tail) but unless the methylation state of this allele in a given individual is known the phenotype cannot be predicted. In contrast to inbred laboratory mouse strains, the human population genetically is highly heterogenous. It is therefore difficult to distinguish between transgenerational epigenetic inheritance and action of several genetic modifiers with additive effects and variable penetrance of mutant alleles (such as seen in A^{vy} and $Axin^{Fu}$ heterozygous carrier mice) observed in human families is likely to be explained by an effect of one or more unlinked modifier genes. The role of epigenetic factors in the etiology of human disease has only recently attracted wide attention and is discussed in detail in other chapters of this book. We are about to witness a breakthrough in the dissection of complex genetic disorders in humans with specific focus on the role of DNA methylation.

When one considers the host defense mechanism and the evolutionary perspective, it is plausible that protection of methylated foreign DNA from demethylation in the primordial germ cells is beneficial. Once a foreign DNA sequence is recognized, methylated, and silenced, it is more economical to preserve this epigenetic information rather than revisit it in every generation. Therefore, to avoid erasure in primordial germ cells, epigenetic marks on foreign DNA would have to differ from epigenetic marks associated with endogenous genes (Hackett et al. 2012). It seems, therefore, a reasonable conjecture that in humans transgenerational epigenetic inheritance could be also detected in association with viral genomes (endogenous retroviruses, hepatitis B, or human immunodeficiency virus) that integrated into the human genomic DNA and may, in principle, be transmitted from one generation to another (Hadchouel et al. 1985; Naumova et al. 1986; Naumova and Kisselev 1990; Tsuei et al. 2002; Wang et al. 2011). In humans, retrotransposon insertions have been implicated in the pathogenesis of a number of cases of single-gene defects (Hancks and Kazazian 2012) but their role in common disease has not been addressed. The modeling of transgenerational epigenetic effects in the human population and their role in evolution are discussed in Chap. 11.

5.3 Epigenetic Memory at Imprinted Regions and Transmission Ratio Distortion

Certain regions in the mammalian genome are imprinted, i.e., carry different DNA methylation marks on the paternal and maternal chromosomes. Imprinted regions are demethylated in primordial germ cells and later remethylated in the germ line in a sex-specific manner, but unlike the gametic marks in the rest of the genome, imprinting marks are maintained in the early embryo (see Chap. 1 for details). This leads to conservation of parent-of-origin-dependent methylation at imprinted regions in embryonic somatic cells and suggests that epigenetic marks other than DNA methylation are present at these regions and guide remethylation of DNA. As a consequence, imprinted regions are more susceptible to DNA methylation resetting errors that occur in the germ line since these errors would not be corrected by global demethylation in the preimplantation period. Therefore imprinted regions are prone to display transgenerational epigenetic inheritance. Because of the pivotal role of imprinted genes in embryonic development, failure to properly reset genomic imprints in the parental germ line would manifest either as embryonic death and infertility or congenital developmental disorders (termed imprinting disorders) in children. Indeed, a proportion of cases with imprinting disorders, including Beckwith–Wiedemann and Prader–Willi syndromes, are associated with abnormal DNA methylation and are consistent with imprint resetting defects in the parental germ line (Reik et al. 1995; Joyce et al. 1997; Buiting et al. 1998; DeBaun et al. 2003; Blik et al. 2006). Moreover, failure to reset genomic imprints has

been proposed as the first hit in a modified version of Knudsen's two-hit model explaining the etiology of embryonic tumors such as retinoblastoma, Wilms tumor, rhabdomyosarcoma, and neuroblastoma (Scrabble et al. 1989, 1990; Sapienza 1991; Naumova and Sapienza 1994). This hypothesis is further supported by the aberrant DNA methylation and expression of imprinted genes in embryonic tumors (Rainier et al. 1993; Ohtani-Fujita et al. 1997; Cui et al. 1998; Feinberg 2000). Furthermore, it has been recently demonstrated that in some individuals, a proportion of mature ejaculated spermatozoa carry somatic instead of gametic DNA methylation patterns in imprinted regions (Marques et al. 2004; Kobayashi et al. 2007, 2009; Marques et al. 2009). Although such abnormal imprinting is more common among infertile men with oligospermia (reduced sperm counts), a number of fertile men also carry abnormal DNA methylation profiles (Marques et al. 2004; Kobayashi et al. 2007, 2009; Marques et al. 2009). A study of DNA methylation imprints in spontaneously aborted fetuses and their fathers demonstrated that the fathers of fetuses with abnormal genomic imprints had abnormal DNA methylation patterns at the same regions in their spermatozoa (Kobayashi et al. 2009). Hence, the fetuses inherited abnormal methylation of imprinted regions from their fathers. These data are the most compelling evidence to date in support of transmission of imprint resetting errors through the germ line in humans.

To fully appreciate the impact of transgenerational epigenetic inheritance in the human population, an estimate of its prevalence is necessary. It could be an extremely rare occurrence associated with specific mutations or it may be a common phenomenon whose impact on the population is reduced by negative selection of gametes or embryos that carry unerased somatic epigenetic marks (Croteau et al. 2001; Naumova et al. 2001).

If failure to reset somatic DNA methylation imprints commonly occurs in a small proportion of germ cells, this will result in parental somatic imprints being transmitted without change to some of the offspring (Naumova et al. 2001; Croteau et al. 2002). For those imprinted regions that harbor genes that are essential for embryonic viability, such an abnormal imprinting pattern would lead to embryonic loss. This loss would occur very early, even before implantation, and would then be detected as deviation from 1:1 Mendelian transmission ratio of alleles at the affected locus. This hypothesis predicts that transgenerational epigenetic inheritance will be detected as non-Mendelian allelic transmission ratios at imprinted genomic regions with preferential transmission of alleles from the grandparent of the same sex as the parent (i.e., alleles from the maternal grandmother or the paternal grandfather) (Naumova et al. 2001). Indeed, grandparental origin-dependent transmission ratio distortion has been found at several imprinted regions in both humans and mice (Naumova et al. 2001; Croteau et al. 2002; Yang et al. 2008). The transmission ratio distortion was in the expected direction, i.e., in most cases the alleles from the maternal grandmother or the paternal grandfather were transmitted to grandchildren with higher probability than the alleles of the paternal grandmother or the maternal grandfather. To test if the molecular mechanism underlying grandparental origin-dependent transmission ratio distortion was indeed related to DNA methylation, the effect of a mutation in the gene for the

major maintenance DNA methyltransferase DNMT1 on transmission ratio distortion was tested in mouse crosses. The transmitting parent was heterozygous for a loss-of-function mutation in the *Dnmt1* gene, which reduced the abundance of *Dnmt1* transcripts in both male and female germ cells and led to reduction in protein concentration. *Dnmt1* mutation in the mother restored Mendelian transmission of maternal alleles at the locus that previously showed transmission ratio distortion among the offspring of wild-type females. Conversely, a *Dnmt1* mutation in the father resulted in deviation from Mendelian transmission ratios of paternal alleles at the same locus (Yang et al. 2008). These data support the role of DNA methylation in the etiology of grandparental origin-dependent transmission ratio distortion. Effects of transmission ratio distortion on genetic analyses are discussed in detail in Chap. 12.

5.4 Genetic Factors and Transgenerational Epigenetic Effects

Mutations in genes that encode factors required for epigenetic reprogramming may modify the likelihood that epigenetic information will be preserved through transmission to the next generation. Among such mutations, those that compromise epigenetic remodeling or maintenance of epigenetic marks in the whole genome will cause infertility and/or embryonic death, as loss of proper epigenetic marks at multiple genomic loci disorganize the finely tuned gene expression ensemble. Striking examples of maternal effects where mutations in the mother result in chaotic DNA methylation in offspring are the targeted mutations of the oocyte-specific isoform of *Dnmt1* RNA, *Dnmt1o*, and the tripartite containing 28 (*Trim28*) gene, which encodes a component of the heterochromatin-inducing complex. Both mutations cause extensive variation in DNA methylation profiles and remarkable phenotypic variation among embryos derived from homozygous mutant females (Howell et al. 2001; Cirio et al. 2008; Toppings et al. 2008; Messerschmidt et al. 2012). Maternal depletion of DNMT1o or TRIM28 renders embryos mosaic for imprinted gene expression which in turn is manifested as variation in embryonic phenotypes ranging from nearly normal appearance to severe malformations (Cirio et al. 2008; Toppings et al. 2008; Messerschmidt et al. 2012) (Fig. 5.3a). The vast majority of embryos from *Dnmt1o*^{-/-} mothers die prenatally at different stages of development independent of their genotypes (Howell et al. 2001; Toppings et al. 2008), therefore analysis of further generations is not possible. However, genetic regulatory variants or mutations that attenuate enzymatic or DNA-binding activity but do not cause catastrophic failure of epigenetic reprogramming are likely to modify the fidelity of reprogramming in the germ line and early embryo.

DNA methylation patterns depend on the genetic background of both the parents and the offspring. Early studies of transgene methylation in mice demonstrated the critical role of the strain harboring the transgene (Weng et al. 1995; Engler et al. 1998). A genetic modifier locus of transgene methylation was mapped to

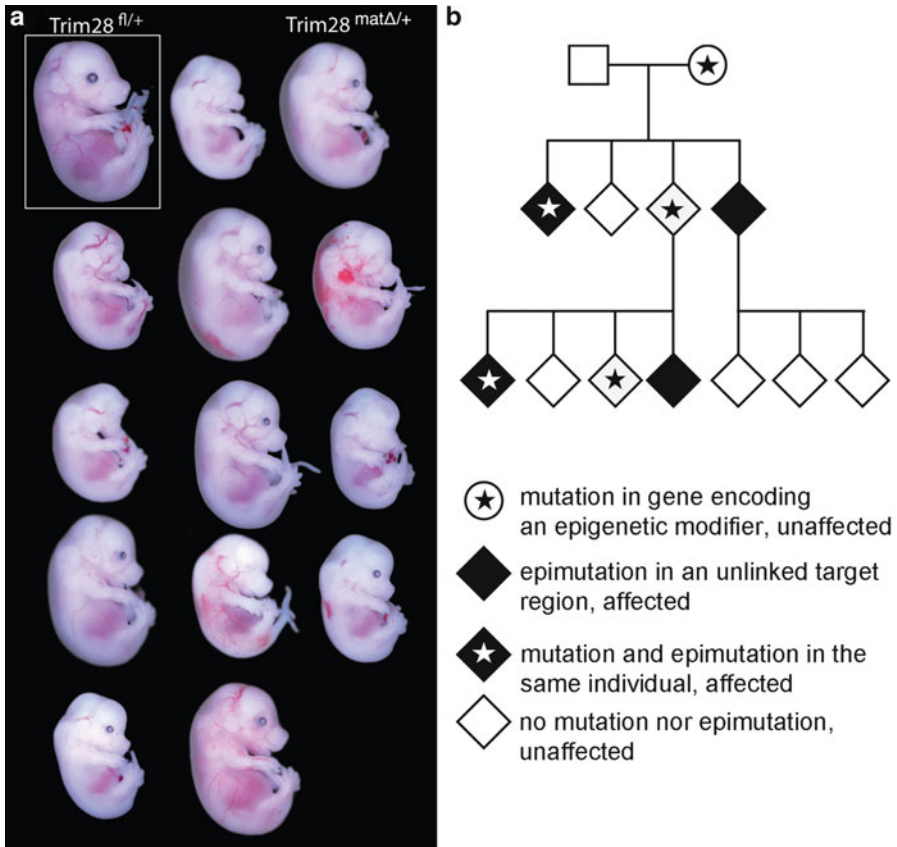


Fig. 5.3 Transgenerational epigenetic effects and complex inheritance patterns. **(a)** Maternal effect mutation in the mouse *Trim28* gene results in high variation in embryo phenotypes (courtesy Dr. D. Messerschmidt; adapted from Messerschmidt et al. *Trim28* Is Required for Epigenetic Stability During Mouse Oocyte to Embryo Transition, *Science* 2012; v. 335, pp. 1499–1502 with permission from The American Association for the Advancement of Science). Mutant E15.5 embryos displaying large array of phenotypes and growth defects compared with a control embryo. **(b)** Mutations in genes encoding epigenetic modifiers or chromosomal translocations that act in the germ line and affect the DNA methylation profiles of their target genes in *trans* segregate independently from the target genomic regions and may or may not be present in the affected individual. Such mutations or chromosomal anomalies exhibit parental effects on phenotype

the distal region of mouse chromosome 4 and was termed strain-specific modifier of transgene methylation 1 (*Ssm1*) (Weng et al. 1995; Engler et al. 1998) (Table 5.1). However, to date, no specific gene has been linked to this methylation activity and the identity of *Ssm1* remains unknown.

Comparison of the *A^{vy}* and *Axin^{Fu}* mouse models uncovered a new dimension to the phenomenon of transgenerational epigenetic inheritance: the effect of genetically defined characteristics of the oocyte cytoplasm on the epigenetic marks

Table 5.1 Genetic modifiers of epigenetic variation and/or epigenetic inheritance

Modifier	Genomic position mouse/human	Gene name and molecular function	Modifying effect on epigenetic phenomena
BAZ1b	Chr. 5 (135.2 Mb)/Chr. 7 (72.8 Mb)	Bromodomain adjacent to zinc finger domain, 1B	Transgene methylation, TEI (Ashe et al. 2008)
DNMT1	Chr. 9 (20.9 Mb)/Chr. 19 (10.2 Mb)	DNA methyltransferase 1	Imprinting (Li et al. 1992), X-chromosome inactivation TRD (Yang 2008), transgene methylation; TEI (Chong et al. 2007b)
DNMT3A	Chr. 12 (3.8 Mb)/Chr. 2 (25.4 Mb)	DNA methyltransferase 3A	Imprinting (Kaneda et al. 2004), TEI (Saferali et al. 2010)
DNMT3L	Chr. 10 (78 Mb)/Chr. 21 (45.6 Mb)	DNA methyltransferase 3L	Imprinting and retrotransposon silencing in the male germ line (Bourc'his and Bestor 2004), meiotic sex chromosome inactivation (Mahadevaiah et al. 2008; Zamudio et al. 2011)
HDAC1	Chr. 4 (129.5 Mb)/Chr. 1 (32.7 Mb)	Histone deacetylase 1	Transgene methylation, TEI (Ashe et al. 2008)
SMARCA5	Chr. 8 (80.7 Mb)/Chr. 4 (144.4 Mb)	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 5	Transgene methylation, TEI (Chong et al. 2007b)
SMCHD1	Chr. 17 (71.3 Mb)/Chr. 18 (2.6 Mb)	SMC hinge domain containing 1	Transgene methylation, TEI (Chong et al. 2007b)
SSM1	Chr4.		Transgene methylation

TEI transgenerational epigenetic inheritance

inherited from the father (Rakyan et al. 2003). Oocytes from inbred C57BL/6 females (C57BL/6 was the background strain in the initial A^{vy} studies) successfully erased the epigenetic marks from the paternal IAP, whereas oocytes from 129/SvJ inbred females, which were used in the *Axin*^{Fu} study, preserved the marks (Rakyan et al. 2003).

To identify genetic modifiers of transgenerational epigenetic inheritance, a screen of mutations that were induced by administering the chemical mutagen *N*-ethyl-*N*-nitrosourea (ENU) to male mice was conducted (Blewitt et al. 2005; Chong et al. 2007b; Ashe et al. 2008). ENU-induced mutations were evaluated for their effects on the epigenetic state and expression of a transgene that contained a green fluorescent protein as a marker of expression and was known for producing phenotypic variegation. Next, each mutation was assessed for its ability to modify transgenerational epigenetic inheritance in the A^{vy} mouse model. The screen identified a number of modifiers of epigenetic inheritance (Table 5.1). Most of those whose functions were previously known turned out to be genes involved in the establishment and regulation of epigenetic states, as one would expect. Genes encoding DNA methyltransferases and genes encoding chromatin modifiers were among those influencing epigenetic variation and transgenerational inheritance (Chong et al. 2007b; Ashe et al. 2008) (Table 5.1). SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 5 (*Smarca5*) encodes a chromatin-remodeling protein that is present in spermatocytes and round spermatids. DNMT1 is the major maintenance DNA methylation enzyme that is responsible for copying the methylation patterns from the original DNA strand to the newly synthesized strand before mitosis. DNMT3L is a member of the methyltransferase family and a cofactor of de novo DNA methyltransferase 3A (Okano et al. 1999; Chen et al. 2005; Ooi et al. 2007). It is therefore possible that in humans, common polymorphisms affecting expression or function of these genes could influence transgenerational epigenetic effects.

5.5 Epigenetic Consequences of Chromosomal Translocations

Heterozygosity for balanced chromosomal translocations interferes with homologous chromosome pairing and synapsis that take place during meiotic prophase. Portions of the translocated chromosomes and their normal homologs that are unable to synapse are silenced by a mechanism termed meiotic silencing of unsynapsed chromatin (MSUC) (reviewed in Burgoyne et al. 2009). The silent chromatin state spreads along the chromosome and silences genes located as far as 10 Mb from the actual translocation breakpoint (Homolka et al. 2007). Interestingly, meiotic silencing is not associated with DNA methylation of gene promoters in the unsynapsed region (Saferali et al. 2010).

In humans, balanced translocations occur in about 0.2 % of live births (Hamerton et al. 1975; Chen et al. 2010; Froslev-Friis et al. 2011). However, their prevalence is

much higher among infertile men and families with recurrent miscarriages (Hassold 1980, 1986; Van Assche et al. 1996). Balanced translocations lead to a wide variety of reproductive outcomes ranging from activation of the mid-pachytene checkpoint mechanism, meiotic arrest, and infertility, problems with chromosome segregation, aneuploid gametes, and embryonic lethality among offspring to completely normal phenotypes. Phenotypic manifestations depend upon the sex of the individual, type of translocation, as well as stochastic factors. It has been hypothesized that the effect of a particular chromosomal translocation on meiotic progression depends upon the functions of genes located within the unsynapsed regions near the translocation breakpoint and silenced by the meiotic silencing mechanism (Burgoyne et al. 2009). If this hypothesis is correct, then translocations occurring near genes that are critical for epigenetic reprogramming during and after meiotic prophase may compromise the resetting of epigenetic marks and thereby increase the probability of epigenetic errors and transgenerational epigenetic inheritance among the offspring of translocation carriers.

This hypothesis was tested in mice that carry a Robertsonian translocation involving chromosomes 8 and 12. The pericentromeric region (translocation breakpoint) of mouse chromosome 12 harbors the DNA methyltransferase 3A (*Dnmt3a*) gene. DNMT3A is a de novo DNA methyltransferase that is essential for the establishment of gametic DNA methylation marks in imprinted regions (Kaneda et al. 2004; Kato et al. 2007). If *Dnmt3a* is silenced during meiotic prophase its deficiency may affect DNA methylation levels throughout the spermatocyte genome. Analysis of DNA methylation of imprinted regions in the sperm of heterozygous translocation carriers showed abnormal DNA methylation at the imprinted *H19* locus on mouse chromosome 7 (Saferali et al. 2010). The *H19* promoter includes a differentially methylated region that is normally methylated in sperm, but unmethylated in oocytes. These data show that chromosomal translocations may cause transgenerational epigenetic effects in *trans*. Thus, depending upon the genetic context of the translocation breakpoint, chromosomal translocations may cause epigenetic disturbances at a specific stage of meiosis and promote transgenerational epigenetic inheritance at loci that segregate independently from the translocation (Fig. 5.3b). This type of inheritance is somewhat reminiscent of the “hit-and-run” mechanism proposed for viral carcinogenesis (Skinner 1976). In the case of balanced chromosomal translocations (similar to genetic modifiers with parental effects) the causal genetic defect segregates independently from its epigenetic target, the gene that is directly responsible for the phenotypic change, and therefore may not be present in the affected individual (Fig. 5.3b). In principle with this type of inheritance, the target genes may be mapped through epigenome-wide association, but not genetic association or linkage mapping approaches. However, the causal genetic defect would only be found if genetic analyses were conducted in parents rather than affected offspring.

To date, however, no specific human pathology has been associated with a parental translocation and such a mode of inheritance. The caveat here is that this specific mechanism has not been researched in genetic studies of human epigenetic disorders. Furthermore, infertility in carriers of balanced translocations affecting the epigenome may be an efficient barrier preventing the birth of children with

epigenetic anomalies. An illustration for this possibility is the increased prevalence of translocations among infertility patients (Bache et al. 2004). The human genome contains a large number of histone genes, some of which are organized into clusters located on chromosomes 1 (1q21 and 1q42) and 6 (6p21–p22). Histones are indispensable for normal genome function in both somatic and meiotic cells. Therefore, one may predict that translocations that occur near these histone gene clusters will cause their meiotic silencing. Transcriptional repression of histone genes starting from the pachytene stage of meiotic prophase will compromise the epigenetic control of the whole genome and lead to meiotic arrest and infertility, or cause transgenerational epigenetic effects in a proportion of live offspring. A survey of balanced translocations in 464 infertile men shows an excess of chromosome 1 translocations, with the most common anomaly involving chromosomal region 1q21 (Bache et al. 2004) near the histone cluster HIST2 in 1q21.2. Hence, abnormal levels of transcription of histone genes during meiosis may be the cause of infertility in 1q21 translocation carriers and potentially increase the risk of epigenetic anomalies among their offspring.

5.6 Environmental Exposures and Transgenerational Epigenetic Effects

The idea that environmental influences (e.g., food supply, exposure to chemicals, or stress) that were experienced by parents or grandparents may define the health (predisposition to diabetes, cardiovascular disease, response to stress, obesity, etc.) of the subsequent generations that were not exposed to such influences is an attractive explanation for complex behavioral and metabolic phenotypes (Pembrey et al. 2006; Crews et al. 2012; Guerrero-Bosagna and Skinner 2012) (reviewed in Latham et al. 2012). This hypothesis has been supported by several studies in different mammals (Anway et al. 2005; Anway and Skinner 2006; Pembrey et al. 2006; Golding et al. 2010; Crews et al. 2012; Guerrero-Bosagna and Skinner 2012; Nilsson et al. 2012). The proposed underlying mechanism is that the environmental stimulus changes the epigenetic profile of certain genes either in the germ line or in the early embryo and the memory of the environmental exposure is thereby transmitted in the form of an epigenetic mark through the parental germ line.

In principle, environmental exposures may cause at least three different types of transgenerational effects.

- (a) Drugs or toxins induce mutations and these mutations in turn cause changes in DNA methylation.
- (b) Environmental factors cause stochastic but stable DNA methylation changes (epimutations) in the absence of genetic mutations. These epimutations are never reversed and will be transmitted unchanged through multiple generations

unless they are lethal or cause infertility (Anway et al. 2005; Chong et al. 2007a; Guerrero-Bosagna and Skinner 2012).

- (c) Environmental factors influence the probability of certain DNA methylation profiles at specific target loci and act in the germ line (Wolff et al. 1998; Cooney et al. 2002; Waterland and Jirtle 2003).

It is also important to keep in mind that living organisms are open systems constantly interacting with the environment, and hence changes in gene expression profiles rapidly occur in response to environmental cues. For example, changes in gene expression are associated with rapid and reversible changes in histone modifications at loci that are essential for the specific response (Holloway et al. 2002; Chinnusamy and Zhu 2009; Koike et al. 2012; Satake and Iwasa 2012). Such environmentally induced epigenetic responses are transient and therefore do not usually cause transgenerational epigenetic inheritance in higher organisms (Chinnusamy and Zhu 2009).

The complexity of phenotypes and the large number of potential target genes make it difficult to dissect the underlying molecular mechanism and directly test the hypothesis of the environmental origins of transgenerational epigenetic effects, particularly in humans. Transmission of epigenetic changes through germ cells has been demonstrated in only a few of these studies (Anway et al. 2005). Moreover, the potential role of new mutations or genetic variants has not been ruled out in many cases (Anway et al. 2005; Kaati et al. 2007; Golding et al. 2010; Borghol et al. 2012). Therefore, strictly speaking, although a number of observations are compatible with transgenerational epigenetic inheritance through the germ line, solid experimental evidence confirming such an inheritance mechanism exists for only a small number of cases. Inbred mouse models, the A^{vy} and *AxinFu* mice, bring insight into the interaction between environment and epigenetic memory. In these mouse models, the phenotype depends on only one mutation, and the role of DNA methylation in the phenotypic manifestation of the genotype has been well established. The A^{vy} mouse was used to investigate whether the levels of folic acid in maternal diet influenced the penetrance of the A^{vy} mutation (Wolff et al. 1998; Cooney et al. 2002; Waterland and Jirtle 2003).

Folic acid is part of the metabolic pathway involved in production of the universal methyl group donor S-adenosyl methionine, which provides methyl groups for DNA and protein methylation. Decreased folic acid levels in the maternal diet during pregnancy reduced the IAP methylation level in the offspring and thereby increased the penetrance of the yellow coat color or the kinked tail phenotype as described above (Wolff et al. 1998; Cooney et al. 2002; Waterland and Jirtle 2003; Waterland et al. 2006). Conversely, folic acid supplementation during pregnancy in the same dams increased the proportion of offspring with high IAP methylation and pseudoagouti coat color or normal tails (Waterland and Jirtle 2003; Waterland et al. 2006). However the diet-induced hypermethylation of the IAP in A^{vy} dams did not accumulate over generations when several generations of females were fed a folic acid-supplemented diet (Waterland et al. 2007). Moreover, the proportion of pseudoagouti mice declined over generations. The most plausible

explanation is that diet-induced DNA methylation is efficient in somatic cells of the F_0 generation, but was actively erased or selectively not maintained in the germ cells of the female embryos in these experiments (Waterland et al. 2007). In summary, environmental agents may induce de novo genetic mutations and heritable epimutations or shift the epigenetic variation in such a way that certain epigenetic variants will have higher chances of appearing among the offspring. In the context of complex traits, environmental exposures may influence the penetrance of phenotypes. By influencing the chances that certain epigenetic variants would occur, environmental exposures may modulate the prevalence of certain phenotypes in the population. The caveat here is that targeted epigenetic modifications of specific genes or regions are not possible at this time. Nonspecific environmental factors such as diet or chemicals affect epigenetic modifications in the whole genome and reducing the risk of disease in one individual may increase the risk in another as shown for the folic acid supplementation and colon cancer development in mice and humans (Knock et al. 2006; Van Guelpen et al. 2006; Lawrance et al. 2009).

Since each of us is born with specific genetic and epigenetic baggage, manipulation of environmental factors seems to be the only practical solution that is available to fight genetic disease. The remote possibility that one day we could improve our phenotypes through targeted modification of epigenetic marks is a very appealing one, if judged by the number of scientific papers on environment and epigenetics. A PubMed search using key words “environment” and “epigenetics” returned a total of 574 citations dating from 1997 to 2012. Of those, 81 % (466) were published in the last 4 years (2009–2012). Therefore, the questions of the degree to which the environment may or may not influence the epigenome; which are the points of epigenetic vulnerability in the individual’s development; as well as the more specific question of how environmental exposures in the parental and grandparental generations influence the phenotypes of their offspring will occupy the minds of researchers for years to come and bring new and exciting developments.

Acknowledgements The author is grateful to Dr. Celia Greenwood, Mary Fujiwara, and Dr. Kenneth Morgan for helpful discussion.

References

- Allen ND, Norris ML, Surani MA (1990) Epigenetic control of transgene expression and imprinting by genotype-specific modifiers. *Cell* 61:853–861
- Amleh A, Xu BZ, Taketo T (2012) Role of sex chromosomes in mammalian female fertility. In: D’Aquino M, Stallone V (eds) *Sex chromosomes: new research*. Nova, New York, NY
- Anway MD, Skinner MK (2006) Epigenetic transgenerational actions of endocrine disruptors. *Endocrinology* 147:S43–S49
- Anway MD, Cupp AS, Uzumcu M, Skinner MK (2005) Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* 308:1466–1469

- Ashe A, Morgan DK, Whitelaw NC, Bruxner TJ, Vickaryous NK, Cox LL, Butterfield NC, Wicking C, Blewitt ME, Wilkins SJ, Anderson GJ, Cox TC, Whitelaw E (2008) A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol* 9:R182
- Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J, Pintacuda G, Sakaguchi A, Sarkies P, Ahmed S, Miska EA (2012) piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 150:88–99
- Bache I, Assche EV, Cingoz S, Bugge M, Tumer Z, Hjorth M, Lundsteen C, Lespinasse J, Winther K, Niebuhr A, Kalscheuer V, Liebaers I, Bonduelle M, Tournaye H, Ayuso C, Barbi G, Blennow E, Bourrouillou G, Brondum-Nielsen K, Bruun-Petersen G, Croquette MF, Dahoun S, Dallapiccola B, Davison V, Delobel B, Duba HC, Duprez L, Ferguson-Smith M, Fitzpatrick DR, Grace E, Hansmann I, Hulten M, Jensen PK, Jonveaux P, Kristoffersson U, Lopez-Pajares I, McGowan-Jordan J, Murken J, Orera M, Parkin T, Passarge E, Ramos C, Rasmussen K, Schempp W, Schubert R, Schwinger E, Shabtai F, Smith K, Stallings R, Stefanova M, Tranebjerg L, Turleau C, van der Hagen CB, Vekemans M, Vokac NK, Wagner K, Wahlstroem J, Zelante L, Tommerup N (2004) An excess of chromosome 1 breakpoints in male infertility. *Eur J Hum Genet* 12:993–1000
- Bestor TH (1998) The host defence function of genomic methylation patterns. *Novartis Found Symp* 214:187–195, discussion 195–189, 228–132
- Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JJ, Arkell R, Whitelaw E (2005) An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A* 102:7629–7634
- Blewitt ME, Vickaryous NK, Paldi A, Koseki H, Whitelaw E (2006) Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet* 2:e49
- Blick J, Terhal P, van den Bogaard MJ, Maas S, Hamel B, Salieb-Beugelaar G, Simon M, Letteboer T, van der Smagt J, Kroes H, Mannens M (2006) Hypomethylation of the H19 gene causes not only Silver-Russell syndrome (SRS) but also isolated asymmetry or an SRS-like phenotype. *Am J Hum Genet* 78:604–614
- Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M, Hertzman C, Power C, Szyf M (2012) Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* 41:62–74
- Bourc'his D, Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431:96–99
- Buiting K, Dittrich B, Gross S, Lich C, Farber C, Buchholz T, Smith E, Reis A, Burger J, Nothen MM, Barth-Witte U, Janssen B, Abeliovich D, Lerer I, van den Ouweland AM, Halley DJ, Schrandt-Stumpel C, Smeets H, Meinecke P, Malcolm S, Gardner A, Lalonde M, Nicholls RD, Friend K et al (1998) Sporadic imprinting defects in Prader-Willi syndrome and Angelman syndrome: implications for imprint-switch models, genetic counseling, and prenatal diagnosis. *Am J Hum Genet* 63:170–180
- Burgoyne PS, Mahadevaiah SK, Turner JM (2009) The consequences of asynapsis for mammalian meiosis. *Nat Rev Genet* 10:207–216
- Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F (2005) Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J Cell Biochem* 95:902–917
- Chen CP, Wu PC, Lin CJ, Su YN, Chern SR, Tsai FJ, Lee CC, Town DD, Chen WL, Chen LF, Lee MS, Pan CW, Wang W (2010) Balanced reciprocal translocations detected at amniocentesis. *Taiwan J Obstet Gynecol* 49:455–467
- Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* 12:133–139
- Chong S, Youngson NA, Whitelaw E (2007a) Heritable germline epimutation is not the same as transgenerational epigenetic inheritance. *Nat Genet* 39:574–575, author reply 575–576
- Chong S, Vickaryous N, Ashe A, Zamudio N, Youngson N, Hemley S, Stopka T, Skoultschi A, Matthews J, Scott HS, de Kretser D, O'Bryan M, Blewitt M, Whitelaw E (2007b) Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat Genet* 39:614–622

- Cirio MC, Martel J, Mann M, Toppings M, Bartolomei M, Trasler J, Chaillet JR (2008) DNA methyltransferase 1o functions during preimplantation development to preclude a profound level of epigenetic variation. *Dev Biol* 324:139–150
- Clark AT (2007) Establishment and differentiation of human embryonic stem cell derived germ cells. *Soc Reprod Fertil Suppl* 63:77–86
- Cooney CA, Dave AA, Wolff GL (2002) Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* 132:2393S–2400S
- Crews D, Gillette R, Scarpino SV, Manikkam M, Savenkova MI, Skinner MK (2012) Epigenetic transgenerational inheritance of altered stress responses. *Proc Natl Acad Sci U S A* 109:9143–9148
- Croteau S, Polychronakos C, Naumova AK (2001) Imprinting defects in mouse embryos: stochastic errors or polymorphic phenotype. *Genesis* 31:11–16
- Croteau S, Freitas-Andrade M, Huang F, Greenwood CMT, Morgan K, Naumova AK (2002) Inheritance patterns of maternal alleles in imprinted regions of the mouse genome at different stages of development. *Mamm Genome* 13:24–29
- Cui H, Horon IL, Ohlsson R, Hamilton SR, Feinberg AP (1998) Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability. *Nat Med* 4:1276–1280
- Davis TL, Trasler JM, Moss SB, Yang GJ, Bartolomei MS (1999) Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* 58:18–28
- Davis TL, Yang GJ, McCarrey JR, Bartolomei MS (2000) The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet* 9:2885–2894
- Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* 13:153–162
- DeBaun MR, Niemitz EL, Feinberg AP (2003) Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *Am J Hum Genet* 72:156–160
- Dickies MM (1962) A new viable yellow mutation in the house mouse. *J Hered* 53:84–86
- Dunn LC, Caspari E (1942) Close linkage between mutations with similar effects. *Proc Natl Acad Sci U S A* 28:205–210
- Engler P, Doglio LT, Bozek G, Storb U (1998) A cis-acting element that directs the activity of the murine methylation modifier locus Ssm1. *Proc Natl Acad Sci U S A* 95:10763–10768
- Feinberg AP (2000) DNA methylation, genomic imprinting and cancer. *Curr Top Microbiol Immunol* 249:87–99
- Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330:622–627
- Fernandez-Gonzalez R, Ramirez MA, Pericuesta E, Calle A, Gutierrez-Adan A (2010) Histone modifications at the blastocyst Axin1(Fu) locus mark the heritability of in vitro culture-induced epigenetic alterations in mice. *Biol Reprod* 83:720–727
- Follette PJ, Laird CD (1992) Estimating the stability of the proposed imprinted state of the fragile-X mutation when transmitted by females. *Hum Genet* 88:335–343
- Froslev-Friis C, Hjort-Pedersen K, Henriques CU, Krogh LN, Garne E (2011) Improved prenatal detection of chromosomal anomalies. *Dan Med Bull* 58:A4293
- Ginsburg M, Snow MH, McLaren A (1990) Primordial germ cells in the mouse embryo during gastrulation. *Development* 110:521–528
- Golding J, Steer C, Pembrey M (2010) Parental and grandparental ages in the autistic spectrum disorders: a birth cohort study. *PLoS One* 5:e9939
- Gomperts M, Garcia-Castro M, Wylie C, Heasman J (1994) Interactions between primordial germ cells play a role in their migration in mouse embryos. *Development* 120:135–141
- Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, Pelisson A, Chambeyron S (2012) piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res* 22:1877–1888
- Grossniklaus U, Kelly B, Ferguson-Smith AC, Pembrey M, Lindquist S (2013) Transgenerational epigenetic inheritance: how important is it? *Nat Rev Genet* 14:228–235

- Guerrero-Bosagna C, Skinner MK (2012) Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Mol Cell Endocrinol* 354:3–8
- Guibert S, Forne T, Weber M (2012) Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res* 22:633–641
- Hackett JA, Reddington JP, Nestor CE, Dunican DS, Branco MR, Reichmann J, Reik W, Surani MA, Adams IR, Meehan RR (2012) Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* 139:3623–3632
- Hackett JA, Sengupta R, Zylitz JJ, Murakami K, Lee C, Down TA, Surani MA (2013) Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 339:448–452
- Hadchouel M, Scotto J, Huret JL, Molinier C, Villa E, Degos F, Brechot C (1985) Presence of HBV DNA in spermatozoa: a possible vertical transmission of HBV via the germ line. *J Med Virol* 16:61–66
- Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 117:15–23
- Hamerton JL, Canning N, Ray M, Smith S (1975) A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin Genet* 8:223–243
- Hancks DC, Kazazian HH Jr (2012) Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 22:191–203
- Hassold TJ (1980) A cytogenetic study of repeated spontaneous abortions. *Am J Hum Genet* 32:723–730
- Hassold T (1986) Chromosome abnormalities in the human reproductive wastage. *Trends Genet* 2:105–110
- Hiura H, Obata Y, Komiyama J, Shirai M, Kono T (2006) Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* 11:353–361
- Holloway AF, Rao S, Shannon MF (2002) Regulation of cytokine gene transcription in the immune system. *Mol Immunol* 38:567–580
- Homolka D, Ivanek R, Capkova J, Jansa P, Forejt J (2007) Chromosomal rearrangement interferes with meiotic X chromosome inactivation. *Genome Res* 17:1431–1437
- Howell C, Bestor T, Ding F, Latham K, Mertineit C, Trasler J, Chaillet J (2001) Genomic imprinting disrupted by a maternal effect mutation in the *Dnmt1* gene. *Cell* 104:829–838
- Howlett SK, Reik W (1991) Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113:119–127
- Joyce JA, Lam WK, Catchpoole DJ, Jenks P, Reik W, Maher ER, Schofield PN (1997) Imprinting of *IGF2* and *H19*: lack of reciprocity in sporadic Beckwith-Wiedemann syndrome. *Hum Mol Genet* 6:1543–1548
- Kaati G, Bygren LO, Pembrey M, Sjöström M (2007) Transgenerational response to nutrition, early life circumstances and longevity. *Eur J Hum Genet* 15:784–790
- Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase *Dnmt3a* in paternal and maternal imprinting. *Nature* 429:900–903
- Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H (2007) Role of the *Dnmt3* family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* 16:2272–2280
- Kearns M, Preis J, McDonald M, Morris C, Whitelaw E (2000) Complex patterns of inheritance of an imprinted murine transgene suggest incomplete germline erasure. *Nucleic Acids Res* 28:3301–3309
- Knock E, Deng L, Wu Q, Leclerc D, Wang XL, Rozen R (2006) Low dietary folate initiates intestinal tumors in mice, with altered expression of G2-M checkpoint regulators polo-like kinase 1 and cell division cycle 25c. *Cancer Res* 66:10349–10356

- Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, Sasaki H, Yaegashi N, Arima T (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16:2542–2551
- Kobayashi H, Hiura H, John RM, Sato A, Otsu E, Kobayashi N, Suzuki R, Suzuki F, Hayashi C, Utsunomiya T, Yaegashi N, Arima T (2009) DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm. *Eur J Hum Genet* 17:1582–1591
- Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS (2012) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338:349–354
- Kuff EL, Lueders KK (1988) The intracisternal A-particle gene family: structure and functional aspects. *Adv Cancer Res* 51:183–276
- Laird CD (1987) Proposed mechanism of inheritance and expression of the human fragile-X syndrome of mental retardation. *Genetics* 117:587–599
- Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35:88–93
- Latham KE, Sapienza C, Engel N (2012) The epigenetic lorax: gene-environment interactions in human health. *Epigenomics* 4:383–402
- Lawrance AK, Deng L, Rozen R (2009) Methylenetetrahydrofolate reductase deficiency and low dietary folate reduce tumorigenesis in *Apc min/+* mice. *Gut* 58:805–811
- Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915–926
- Lim JP, Brunet A (2013) Bridging the transgenerational gap with epigenetic memory. *Trends Genet* 29:176–186
- Lucifero D, Mann MR, Bartolomei MS, Trasler JM (2004) Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum Mol Genet* 13:839–849
- Mahadevaiah SK, Bourc'his D, de Rooij DG, Bestor TH, Turner JM, Burgoyne PS (2008) Extensive meiotic asynapsis in mice antagonises meiotic silencing of unsynapsed chromatin and consequently disrupts meiotic sex chromosome inactivation. *J Cell Biol* 182:263–276
- Maksakova IA, Mager DL, Reiss D (2008) Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell Mol Life Sci* 65:3329–3347
- Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. *Lancet* 363:1700–1702
- Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, Sousa M (2009) Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil Steril* 1:1
- Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* 130:851–862
- Matzke MA, Matzke AJ (1998) Epigenetic silencing of plant transgenes as a consequence of diverse cellular defence responses. *Cell Mol Life Sci* 54:94–103
- McGowan R, Campbell R, Peterson A, Sapienza C (1989) Cellular mosaicism in the methylation and expression of hemizygous loci in the mouse. *Genes Dev* 3:1669–1676
- Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB (2012) Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* 335:1499–1502
- Monk M (1995) Epigenetic programming of differential gene expression in development and evolution. *Dev Genet* 17:188–197
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23:314–318
- Naumova AK, Kisselev LL (1990) Biological consequences of interactions between hepatitis B virus and human nonhepatic cellular genomes. *Biomed Sci* 1:233–238
- Naumova A, Sapienza C (1994) The genetics of retinoblastoma, revisited. *Am J Hum Genet* 54:264–273

- Naumova AK, Korenev VI, Leonov BV, Tsibinogin VV, Kiselev LL (1986) DNA of the hepatitis B virus in human generative cells. *Genetika* 22:166–168
- Naumova AK, Greenwood CM, Morgan K (2001) Imprinting and deviation from Mendelian transmission ratios. *Genome* 44:311–320
- Nelson VR, Nadeau JH (2010) Transgenerational genetic effects. *Epigenomics* 2:797–806
- Nilsson E, Larsen G, Manikkam M, Guerrero-Bosagna C, Savenkova MI, Skinner MK (2012) Environmentally induced epigenetic transgenerational inheritance of ovarian disease. *PLoS One* 7:e36129
- Obata Y, Kono T (2002) Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. *J Biol Chem* 277:5285–5289
- Ohtani-Fujita N, Dryja TP, Rapaport JM, Fujita T, Matsumura S, Ozasa K (1997) Hypermethylation in the retinoblastoma gene is associated with unilateral, sporadic retinoblastoma. *Cancer Genet Cytogenet* 98:43–49
- Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448:714–717
- Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjöström M, Golding J, Team AS (2006) Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* 14:159–166
- Perry WL, Copeland NG, Jenkins NA (1994) The molecular basis for dominant yellow agouti coat color mutations. *BioEssays* 16:705–707
- Pickard B, Dean W, Engemann S, Bergmann K, Fuermann M, Jung M, Reis A, Allen N, Reik W, Walter J (2001) Epigenetic targeting in the mouse zygote marks DNA for later methylation: a mechanism for maternal effects in development. *Mech Dev* 103:35–47
- Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP (1993) Relaxation of imprinted genes in human cancer. *Nature* 362:747–749
- Rakyan VK, Blewitt ME, Druker R, Preis JI, Whitelaw E (2002) Metastable epialleles in mammals. *Trends Genet* 18:348–351
- Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV, Whitelaw E (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci U S A* 100:2538–2543
- Reed SC (1937) The inheritance and expression of fused, a new mutation in the house mouse. *Genetics* 22:1–13
- Reik W, Collick A, Norris ML, Barton SC, Surani MA (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* 328:248–251
- Reik W, Brown KW, Schneid H, Le Bouc Y, Bickmore W, Maher ER (1995) Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by altered imprinting pattern in the IGF2-H19 domain. *Hum Mol Genet* 4:2379–2385
- Saferali A, Berlivet S, Schimenti J, Bartolomei MS, Taketo T, Naumova AK (2010) Defective imprint resetting in carriers of Robertsonian translocation Rb (8.12). *Mamm Genome* 21:377–387
- Sapienza C (1991) Genome imprinting and carcinogenesis. *Biochim Biophys Acta* 1072:51–61
- Sapienza C, Peterson AC, Rossant J, Balling R (1987) Degree of methylation of transgenes is dependent on gamete of origin. *Nature* 328:251–254
- Satake A, Iwasa Y (2012) A stochastic model of chromatin modification: cell population coding of winter memory in plants. *J Theor Biol* 302:6–17
- Scrabble H, Caveness W, Ghavimi F, Lovell M, Morgan K, Sapienza C (1989) A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. *Proc Natl Acad Sci U S A* 86:7480–7484
- Scrabble HJ, Sapienza C, Caveness WK (1990) Genetic and epigenetic losses of heterozygosity in cancer predisposition and progression. *Adv Cancer Res* 54:25–62

- Skinner GR (1976) Transformation of primary hamster embryo fibroblasts by type 2 simplex virus: evidence for a “hit and run” mechanism. *Br J Exp Pathol* 57:361–376
- Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat Genet* 43:811–814
- Swain JL, Stewart TA, Leder P (1987) Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. *Cell* 50:719–727
- Szabo PE, Mann JR (1995) Biallelic expression of imprinted genes in the mouse germ line: implications for erasure, establishment, and mechanisms of genomic imprinting. *Genes Dev* 9:1857–1868
- Toppings M, Castro C, Mills PH, Reinhart B, Schatten G, Ahrens ET, Chaillet JR, Trasler JM (2008) Profound phenotypic variation among mice deficient in the maintenance of genomic imprints. *Hum Reprod* 23:807–818
- Trasler JM (2009) Epigenetics in spermatogenesis. *Mol Cell Endocrinol* 306:33–36
- Tsuei DJ, Chang MH, Chen PJ, Hsu TY, Ni YH (2002) Characterization of integration patterns and flanking cellular sequences of hepatitis B virus in childhood hepatocellular carcinomas. *J Med Virol* 68:513–521
- Valenza-Schaerly P, Pickard B, Walter J, Jung M, Pourcel L, Reik W, Gauguier D, Vergnaud G, Pourcel C (2001) A dominant modifier of transgene methylation is mapped by QTL analysis to mouse chromosome 13. *Genome Res* 11:382–388
- Van Assche E, Bonduelle M, Tournaye H, Joris H, Verheyen G, Devroey P, Van Steirteghem A, Liebaers I (1996) Cytogenetics of infertile men. *Hum Reprod* 11(Suppl 4):1–24, discussion 25–26
- Van Guelpen B, Hultdin J, Johansson I, Hallmans G, Stenling R, Riboli E, Winkvist A, Palmqvist R (2006) Low folate levels may protect against colorectal cancer. *Gut* 55:1461–1466
- Vasicek TJ, Zeng L, Guan XJ, Zhang T, Costantini F, Tilghman SM (1997) Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics* 147:777–786
- Wang D, Li LB, Hou ZW, Kang XJ, Xie QD, Yu XJ, Ma MF, Ma BL, Wang ZS, Lei Y, Huang TH (2011) The integrated HIV-1 provirus in patient sperm chromosome and its transfer into the early embryo by fertilization. *PLoS One* 6:e28586
- Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23:5293–5300
- Waterland RA, Dolinoy DC, Lin JR, Smith CA, Shi X, Tahiliani KG (2006) Maternal methyl supplements increase offspring DNA methylation at Axin Fused. *Genesis* 44:401–406
- Waterland RA, Travisano M, Tahiliani KG (2007) Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. *FASEB J* 21:3380–3385
- Weng A, Magnuson T, Storb U (1995) Strain-specific transgene methylation occurs early in mouse development and can be recapitulated in embryonic stem cells. *Development* 121:2853–2859
- Wolff GL (1965) Body composition and coat color correlation in different phenotypes of “viable yellow” mice. *Science* 147:1145–1147
- Wolff GL (1978) Influence of maternal phenotype on metabolic differentiation of agouti locus mutants in the mouse. *Genetics* 88:529–539
- Wolff GL, Kodell RL, Moore SR, Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J* 12:949–957
- Yang L, Andrade MF, Labialle S, Moussette S, Geneau G, Sinnett D, Belisle A, Greenwood CM, Naumova AK (2008) Parental effect of DNA (cytosine-5) methyltransferase 1 on grandparental-origin-dependent transmission ratio distortion in mouse crosses and human families. *Genetics* 178:35–45
- Zamudio NM, Scott HS, Wolski K, Lo CY, Law C, Leong D, Kinkel SA, Chong S, Jolley D, Smyth GK, de Kretser D, Whitelaw E, O’Byrne MK (2011) DNMT3L is a regulator of X chromosome compaction and post-meiotic gene transcription. *PLoS One* 6:e18276

Chapter 6

Autosomal Monoallelic Expression

Virginia Savova and Alexander A. Gimelbrant

Abstract In mammals, relative expression of the maternal and paternal alleles of many genes is controlled by three types of epigenetic phenomena: X chromosome inactivation, imprinting, and mitotically stable autosomal monoallelic expression (MAE). MAE imposes a mitotically stable allelic imbalance in the expression of a significant fraction of human autosomal genes. Cells in the same individual make independent choices of active and inactive alleles, leading to remarkable epigenetic diversity between otherwise identical clonal lineages. Genes subject to MAE play critical roles in a variety of major disorders, including schizophrenia, Alzheimer's disease, and cancer. In this chapter, we review the current state of understanding of MAE biology, and assess various implications of MAE for analysis of genotype–phenotype relationship.

6.1 Biology of Autosomal Monoallelic Expression

A variety of genetic and epigenetic factors affect the relative expression levels of the two copies of each given gene in diploid cells. In addition to genetic variation in regulatory regions that affects allele-specific expression (Cowles et al. 2002; Yan et al. 2002), there are at least three major kinds of non-Mendelian, epigenetic

V. Savova

Department of Cancer Biology, Dana-Farber Cancer Institute, 450 Brookline Ave,
Boston, MA 02115, USA

Department of Genetics, Harvard Medical School, 450 Brookline Ave, Boston,
MA 02115, USA

A.A. Gimelbrant (✉)

Department of Genetics, Harvard Medical School, 450 Brookline Ave, Boston,
MA 02115, USA

Department of Cancer Biology, Dana-Farber Cancer Institute, Smith Building,
Room 922 A, 450 Brookline Ave., Boston, MA 02115, USA
e-mail: Gimelbrant@mail.dfci.harvard.edu

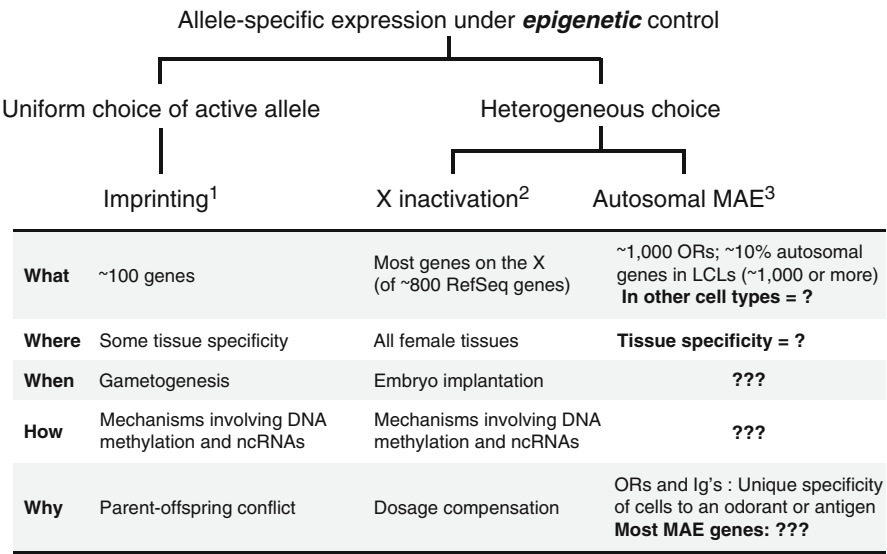


Fig. 6.1 Autosomal monoallelic expression in the context of epigenetic mechanisms controlling allele-specific expression in mammalian cells. Main mechanistic, developmental, and functional questions for the three major mechanisms affecting allelic expression imbalance. *Notes:* 1 = reviewed in Abramowitz and Bartolomei (2012) and Kelsey and Bartolomei (2012); 2 = reviewed in Augui et al. (2011) and Jeon et al. (2012); 3 = see references throughout this chapter

phenomena that control allele-specific expression in mammals (see Fig. 6.1). One is the X chromosome inactivation: During the development of female embryos, around the time of implantation, about half of the cells choose to inactivate the maternal X, and the rest inactivate the paternal X, affecting most of the X-linked genes (Berletch et al. 2010; Carrel and Willard 2005; Lyon 1961; Yang et al. 2010). Another is imprinting: Genes such as *IGF2* and *H19* are expressed either from paternal or from maternal allele (reviewed in Glaser et al. 2006). These two epigenetic mechanisms are reviewed in the companion Chaps. 1, 3, and 5 in this volume.

In addition, a significant fraction of mammalian genes are subject to *autosomal monoallelic expression* (MAE). MAE is observed in olfactory receptor genes (Chess et al. 1994), as well as genes coding for immunoglobulins and some cytokines. Using genome-wide analyses of allele-specific expression, we and others have added a surprisingly large number of the autosomal genes in human and mouse to the MAE class, including genes implicated in a number of major human diseases, such as Alzheimer’s disease (*APP*) and cancer (*DAPK1*) (Gimelbrant et al. 2007; Jeffries et al. 2012; Li et al. 2012; Zwemer et al. 2012). MAE affects about 10 % of ~4,000 tested genes in human lymphoblastoid cells and about 15 % of more than 1,300 assessed genes in equivalent mouse cells. Note that this count excludes olfactory receptor genes which by themselves constitute about 5 % of mammalian protein-coding genes. There is also evidence that ribosomal DNA gene clusters are subject to mitotically stable MAE (Schlesinger et al. 2009). Overall, of

the three epigenetic mechanisms (depicted in Fig. 6.1) MAE affects by far the greatest number of genes.

MAE is an “autosomal analog of X inactivation” in the sense that it creates epigenetic mosaicism: Otherwise identical cells in the same individual have different “epigenotypes” such that some cells express the maternal copy of a gene, and other cells express the paternal copy of that gene. If the two alleles of the gene are functionally distinct, this can result in emergence of subpopulations of cells with different functional properties, such as differential responsiveness to lipopolysaccharides in lymphocytes of mice heterozygous for a mutant variant of the *Tlr4* gene (Pereira et al. 2003). Since X chromosome inactivation is much better known, we will structure this discussion by comparing and contrasting MAE and X inactivation.

Similarly to X inactivation, once the choice of the active allele is made, it is maintained in a mitotically stable manner. For example, in one experiment, mouse cells maintained near-complete silencing of one copy of p120 catenin after a year in continuous culture (Gimelbrant et al. 2005). More generally, since most systematic assessments of MAE were performed on cell populations grown from a single cell to 10^7 – 10^8 cells (Gimelbrant et al. 2007; Zwemer et al. 2012), we can conclude with confidence that the allelic choices genome-wide have been maintained through dozens of cell divisions.

One important consequence of these stable differences between clonal lineages is that in a heterogeneous, polyclonal samples MAE could pass undetected (see Fig. 6.2). Not coincidentally, MAE was discovered by analyzing specific clonal cell populations of immune cells (Bix and Locksley 1998; Hollander et al. 1998; Pernis et al. 1965), or in the context of analysis of individual cells, such as olfactory sensory neurons (Chess et al. 1994), and kept resurfacing when single-cell or clonal population approaches were used (Takizawa et al. 2008). Strictly speaking, if a gene with MAE is not mitotically stable and allows switching of expression between alleles upon mitosis, it would remain undetected in an analysis of allele-specific expression, even in a single-cell clone. We have seen no evidence of such instability in individual genes tested using short clone expansion times (analyzed after expanding to few hundred cells) or FISH analysis of individual cells from a population showing biallelic expression of a given gene (Gimelbrant et al. 2005, 2007), but it remains possible that genes that show such labile patterns do exist.

Unlike X inactivation, MAE is observed both in male and female cells. In genes that show MAE, allelic bias is often very strong: complete silencing of one allele for olfactory receptors and immunoglobulins, and tenfold or greater bias in allelic expression for many other autosomal genes (Gimelbrant et al. 2007). At the same time, MAE appears more variable than X inactivation. While X inactivation is ubiquitous (i.e., occurs in all somatic cells of a postimplantation female embryo), currently, there is no definitive knowledge of tissue specificity of MAE, as even clones of the same cell type are different from each other. However, small patches of placenta and clonal fibroblasts have similar fraction of MAE genes (Gimelbrant et al. 2007), and analysis of individual MAE genes in multiple clones

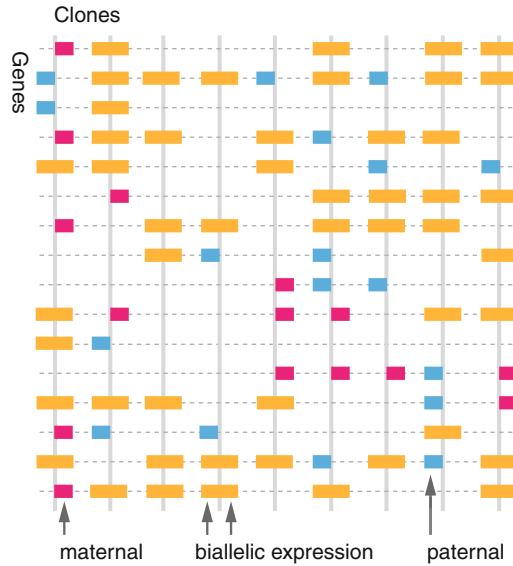


Fig. 6.2 Clonal cells show unique pattern of monoallelic expression. Allelic choice for MAE genes in human lymphoblastoid cells. Schematic representation of maternal, paternal, or biallelic expression of MAE genes on chromosome 18 (based on the data from Gimelbrant et al. 2007). Vertical lines correspond to individual clones; dotted horizontal lines mark genes. Gene marker to the right of vertical line—maternal expression; to the left—paternal expression; present on both sides—biallelic expression. Nine clones were derived from three individuals. Note the lack of coordination of allelic choice along the chromosome in the same clone, and independent choice of allelic state of the same gene between clones

shows that a gene could be subject to MAE in one tissue and not subject in another (Gimelbrant et al. 2005).

While MAE is obligatory for olfactory receptor genes (Serizawa et al. 2000) and in the allelic exclusion of immunoglobulin genes (Jung et al. 2006), for most other autosomal MAE genes, MAE is not obligatory: in mouse and human lymphocytes, more than 80 % of MAE genes were biallelic in at least one assessed clone (Gimelbrant et al. 2005, 2007). This somewhat resembles variability among genes that escape X inactivation (Carrel and Willard 2005), though it is unlikely that similar mechanisms are involved. The presence of two active alleles vs. one active allele in clones of otherwise similar cells resulted in (about twofold) higher expression levels at least for a small number of assessed genes (Gimelbrant et al. 2005, 2007). In this connection, it is worth noting that the temporal changes in *Nanog* RNA levels during differentiation correlate with its allelic state: MAE corresponds to lower levels, and higher RNA levels to biallelic expression (Miyazari and Torres-Padilla 2012). A systematic study is needed to establish whether that is a general rule.

6.2 Open Questions in MAE Biology

Our growing appreciation of the prevalence of MAE only underscores how little we know about its biology. Here we summarize unresolved questions (see also Fig. 6.1).

MAE was directly observed *in vivo* using fluorescent protein reporters in interleukin-4 (Hu-Li et al. 2001) and olfactory receptor genes (Serizawa et al. 2000), and fluorescent *in situ* hybridization in olfactory epithelium (Chess et al. 1994) and in human peripheral blood lymphocytes for several genes initially detected as MAE in lymphoblastoid cells (Gimelbrant et al. 2007). However, MAE clonality makes its assessment on a genome-wide scale very challenging; thus the only large-scale sets of data are collected in clonal cell lines *in vitro*, primarily lymphoblastoid cells. The limited number of clones thus analyzed is insufficient to generate a complete catalog of MAE genes in lymphoblastoid cells, and little is known about the prevalence of MAE in other cell types. Virtually nothing is known about the establishment of MAE during development and differentiation.

Mechanistically, allelic choice has been linked to changes in chromatin states in two special cases: olfactory receptor gene choice (Magklara et al. 2011) and immunoglobulin-kappa gene rearrangement (Farago et al. 2012). In contrast, for hundreds of other autosomal MAE genes, no molecular features have so far been reported to be strongly associated with the establishment and maintenance of allelic choice, even though intriguing correlations have been observed: (1) MAE genes are more likely to be located close to recombination hot spots (Necsulea et al. 2009) and (2) near clusters of related genes (Gimelbrant and Chess 2006) and (3) promoters of MAE genes are enriched in a statistically significant manner with chromatin modifications associated with both open chromatin (histone H3 Lys-4 methylation) and inactive chromatin (histone H3 Lys-27 methylation) (Bock et al. 2009). Non-imprinted allele-specific DNA methylation had been reported in lymphoblasts (Zhang et al. 2009); however, since cell culture might severely affect DNA methylation levels (Antequera et al. 1990), making lymphoblasts potentially unreliable in this respect (Saferali et al. 2010), the role of DNA methylation in MAE remains unclear.

In a marked contrast with X inactivation, there is no chromosome-wide coordination in the choice of the active and inactive alleles. Instead, genes subject to MAE are interspersed with biallelic genes throughout the genome, with each locus' allelic choice apparently independent of any other locus (see Fig. 6.2). A parsimonious explanation posits that allelic choice involves separate, independent regulatory regions in proximity of each autosomal MAE locus. These hypothetical sequences can be conceptualized as being functionally similar to X inactivation center (Augui et al. 2011), but affecting a small region rather than the whole chromosome.

Finally, there is no satisfactory general explanation of the biological function of MAE. In specialized cases of olfactory receptors and immunoglobulins, plausible functional accounts exist: these are the receptors that impose a unique and precise specificity onto otherwise identical cells (such as T- or B-lymphocytes, or sensory

neurons), which could be impaired by co-expression of the other allele. There is no such intuitively appealing hypothesis with regard to the majority of MAE genes, such as interleukins or amyloid precursor protein. One possible hint is provided by the observation that among MAE genes, there is a strong overrepresentation of genes coding for cell-surface proteins (Gimelbrant et al. 2007). Between dosage variation (for non-obligatory MAE genes) and independent choice of allele at each of many hundreds of loci, MAE can serve as a combinatorial mechanism providing a unique signature to any number of cells or clonal lineages at their interface to extracellular environment. This epigenetic diversity might be functionally important in the development of complex organs (cf. role played by DSCAM gene in *Drosophila*: due to extensive alternative splicing, it can encode thousands of isoforms, whose homophilic repulsion is the basis of neuronal self-recognition and self-avoidance) (reviewed in Hattori et al. 2008). We propose that such diversity can play a crucial biological role by providing “passive immunity”: an epigenetic mosaic of clones, by virtue of presenting subtly distinct combinations of surface proteins to infectious agents, can have different levels of susceptibility rather than form a uniformly susceptible “monoculture.” Finally, MAE might be a nonadaptive consequence of some other property of particular genomic areas. To perform a crucial experiment determining MAE function, one would need to identify mechanisms involved in the establishment and maintenance of the MAE, eliminate MAE in a model organism, and assess the impact of that manipulation in the organism’s development and function.

6.3 Autosomal Monoallelic Expression, Clonality, and Analysis of Genotype–Phenotype Relationship

Sequence-based variation in gene expression is a key driver of disease risk. Identification of sequence variants regulating expression in *cis* is a major effort in genetics of complex traits (Gilad et al. 2008). Whatever proves to be the ultimate explanation of the MAE function, existence of MAE in mammalian cells introduces a number of challenges into analysis and interpretation of a relationship between genotype and phenotype, both in the context of efforts to understand the biology of gene regulation and in the context of personalized medicine.

The central issue is epigenetic heterogeneity introduced by MAE—the fact that clonal lineages which are “siblings” of each other can be quite distinct in the spectrum of alleles of MAE genes that are chosen and are maintained over long periods of time (Fig. 6.3). This epigenetic heterogeneity means that knowledge of the average expression state in a polyclonal tissue as a whole does not translate into precise information about the state of individual cells. Conversely, the state of a particular clonal lineage may not translate into complete information about overall level of expression or allelic bias in a given cell type, even when assessing cells with exactly the same genotype (e.g., cells from the same individual).

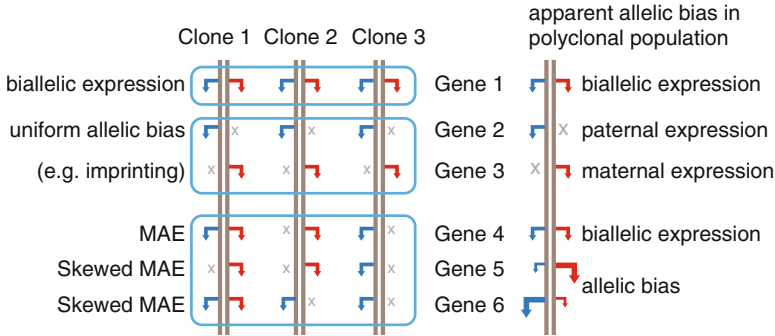


Fig. 6.3 Relationship between allele-specific expression in individual clones and in polyclonal tissue. Schematic depiction of clone-specific allelic choice for multiple genes in three hypothetical clonal cell subpopulations in one individual, and the resulting allelic (im)balance in the polyclonal tissue. Each clone is represented by a pair of chromosomes (shown as vertical lines). Angled arrows—active transcription; left-side arrow—paternal expression; right-side arrow—maternal expression; cross—transcriptional silencing

This is not a radically novel problem: indeed, it is quite similar to pronounced differences between cells of the same type that X inactivation can lead to. The heterogeneity between individual cells that combine into cell type average expression values is particularly important when considering cell-autonomous functions, and in particular, tumor initiation. In case of X inactivation, a remarkable example can be found in women heterozygous for a loss-of-function variant of *FOXP3* gene (Zuo et al. 2007). This gene, encoding a transcription factor with tumor suppressor activity, is X-linked in both humans and mice, and heterozygous knockout mice developed mammary tumors at a high rate. When clinical samples of breast cancer in heterozygous women were analyzed, the tumors nearly always arose from cells that had inactivated the X chromosome with the functional copy of the gene resulting in a functional loss of *FOXP3*. In other words, different epigenotypes (with respect to the activity of one or the other copy of *FOXP3*) led to dramatically different outcomes in cell fate.

Below we discuss several situations in which one might want to be aware of MAE while dissecting the interaction between genotype and phenotype.

Analysis of genetic and epigenetic architecture of allele-specific expression. There are several potential biological sources of deviation from a 50:50 allelic expression ratio: a genetic effect of *cis*-regulatory variation, and several epigenetic/developmental effects: a parent-of-origin effect (imprinting), a sampling effect of clonality, and primary or secondary skewing of initial establishment of the allelic choice. While the effects of those on X-linked genes have been discussed (Chadwick and Willard 2005; Heard et al. 1997; Wang et al. 2010), the clonality effects are not routinely taken into consideration for autosomal genes. Note that in the following discussion of the clonality effects we completely leave aside any questions of technical noise or artifactual systematic bias introduced during measurement (such as preferential amplification or detection of one allele over the other); we focus on biological factors affecting allelic bias as it “truly is” in the sample.

Collection sampling effect. The process of sampling biological material can lead to overrepresentation of particular clonal lineage. For example, lymphoblastoid cell lines, which are commonly used in expression quantitative trait loci (eQTL) studies of multiple individuals (Dixon et al. 2007; Lee et al. 2013), can be monoclonal or oligoclonal in 20 % of cases (Pastinen et al. 2004; Plagnol et al. 2008). For MAE genes, the allelic bias in these samples is determined by the choice of alleles in the dominant clones and may not be representative of the cell type as a whole. This could be particularly salient for studies designed to reliably distinguish *cis*-regulatory effects from *trans* effects by measuring allele-specific expression at heterozygous loci (Hull et al. 2007; Pickrell et al. 2010).

Even samples that are perfectly polyclonal and representative of a given individual can be affected by the *sampling effect* caused by the small number of cells at the moment of commitment to the active allele choice. Depending on the developmental time when the active allele is chosen, the number of cells involved might be fairly small [at the time of X inactivation, there are just about 15 cells that will give rise to the entire hematopoietic system (Tonon et al. 1998)], resulting in large skewing in a significant fraction of individuals.

Moreover, an observed allelic bias in polyclonal samples can result from primary and secondary allelic skewing. *Primary skewing* signifies the idea of “unfair coin” used in deciding which allele will be active in a given lineage. In case of X inactivation in mice, this is determined by which variant of *Xce* locus is present on each of the X chromosomes (Chadwick et al. 2006; Chadwick and Willard 2005; Heard et al. 1997; Plenge et al. 2000). For example, in Cast/Ei 129 F1 cross, only about 20 % of cells will have X^{129} active and X^{Cast} inactive. There is evidence that skewing affects MAE genes, at least in the mouse cells: for multiple genes, one allele is much more likely than the other to be found the only one expressed (Zwemer et al. 2012). While such primary allelic skewing is presumably genetically driven, its consequences differ from *cis*-acting variation, which causes the uniform allelic skewing present in each cell. In case of MAE, average allelic imbalance masks a fraction of cells that may be biallelic, or skewed in the other direction. *Secondary clonal skewing* can result from preferential survival of a clone (Plenge et al. 2002) that could be related to its functional properties, or could be due to stochastic events during developmental bottlenecks. By studying mother–neonate pairs Bolduc et al. (2008) show that X inactivation skewing is both present at significant rates in humans (8–28 % dependent on age and tissue type) and not heritable, suggesting secondary skewing as the dominant cause.

Analysis of total expression level and its relationship to function. Because of the propensity of MAE genes to show stably biallelic expression in a significant fraction of cells, and biallelic expression typically corresponding to higher level of transcript than monoallelic expression, the clonal composition of a sample affects the overall expression level of the gene, apart from having an effect on allelic bias. In a tissue, the average level of expression of an MAE gene is determined by its RNA levels in clones with biallelic, paternal, or maternal monoallelic expression, as well as the fractional representation of each type of

clone in the general population of cells. Thus the expression data from a particular clonal sample is not necessarily representative of the cell type as a polyclonal whole, and vice versa. This should be taken into account when systematically comparing samples of different clonality, such as tumors and matched normal tissue.

This has bearing on the important question of identification of potential oncogenes. Overexpression in tumor samples compared to normal tissue even without evidence for DNA amplification is considered a principal form of evidence for oncogene identification (Ko et al. 2003; Santarius et al. 2010). A common strategy for uncovering oncogenic mutations in non-blood cancers is to obtain genotype and expression data from tumors and matched blood samples from the same patients (Li et al. 2013), or large-scale transcriptional profiling using tumor samples from patients and normal tissue samples from healthy individuals (Gordon et al. 2005). Clonal variability of expression levels of MAE genes is a likely confounding factor. A clonal outgrowth originating from a cell that happened to have a gene in question biallelically expressed will exhibit overexpression compared to normal tissue. For example, if 30 % of the cells in the normal tissue have biallelic expression for a given MAE gene, we can expect that approximately 30 % of tumors originating in cells of that type will show evidence of overexpression in that gene, even if the gene in question has no effect on tumor initiation or progression.

At least in some cases, the switch between monoallelic and biallelic expression and associated change in effective gene dosage do play a functional role—see an example with *Nanog* gene during differentiation (Miyanari and Torres-Padilla 2012). Detailed information on the clonal composition and primary allelic skewing in the normal tissue across individuals would allow testing of the hypothesis that such composition plays a role in disease etiology.

Analysis of nongenetic variability. MAE is likely to be an important contributing mechanism in nongenetic variability which plays a role in a range of conditions. One classic case of somatic functional mosaicism is field cancerization (Slaughter et al. 1953): emergence of multiple primary tumors in a limited area in close proximity to each other. In some cases this phenomenon can be ascribed to changes brought by somatic mutations [e.g., as a result of UV-damage to stromal cells (Ratushny et al. 2012)]. Such changes can also be epigenetically driven (Hu et al. 2012). Clonal nature of MAE makes it a likely candidate mechanism for such local, mosaic effects. Intriguingly, a recent work (Kreso et al. 2012) reported that epigenetic clonal heterogeneity is maintained in colon cancer in a way that affects response to chemotherapy for individual clones and contributes to development of chemotherapy resistance. MAE, with its capacity to create remarkable epigenetic heterogeneity and maintain it over multiple cell divisions, is a promising candidate for the underlying mechanism. This also holds a potential for MAE-focused epigenetic treatments to prevent tumor initiation, as well as reduce drug resistance in chemotherapy.

Haploinsufficiency. A final, particularly intriguing relationship is the one between MAE and haploinsufficiency in those cases where one wild-type allele in a heterozygote is insufficient for normal function. Haploinsufficiency has been reported in neurodevelopmental disorders such as Williams syndrome (Meng et al. 1998) and language delays (Lamb et al. 2012); it also affects a significant number of genes with tumor suppressor activity (Berger and Pandolfi 2011). MAE raises the possibility that haploinsufficiency comes in two distinct types. In the first, uniform type, every cell in affected tissue has similar, insufficient level of expression of the non-MAE gene in question. In the second type, MAE would lead to nonuniformity in cells forming the affected tissue, with some cell subpopulations completely lacking a functional copy of the gene and other cells completely normal. The latter cases would be more likely to result in variable phenotype due to variation in clonal composition and potential rescue through allelic skewing [somewhat reminiscent of hemizygous mutations in the *FMR* gene leading to variable phenotype of fragile X syndrome in female (Kirchgessner et al. 1995)]. The MAE-based type of haploinsufficiency may be amenable to a different set of therapeutic strategies than the uniform type, such as finding a way to reverse the epigenetic allelic choice or derepress the silent allele.

6.4 Conclusion

Autosomal MAE is a widespread epigenetic phenomenon that affects more than 10 % of human and mouse genes. Many mechanistic questions related to MAE are open. A major technical bottleneck in addressing these questions is the clonal nature of MAE: like X inactivation, MAE is masked in polyclonal samples. Depending on the clonal composition of a biological sample, MAE can contribute to apparent allelic expression bias, as well as to noticeable differences in expression levels. While it can present analytical challenges in a number of research strategies and personalized medicine approaches, MAE also holds promise for epigenetic manipulation of functional properties of cells and cell populations.

Acknowledgements We thank Anwesha Nag for figure design, and Sebastien Vigneau and Gimelbrant lab members for stimulating discussions.

References

- Abramowitz LK, Bartolomei MS (2012) Genomic imprinting: recognition and marking of imprinted loci. *Curr Opin Genet Dev* 22:72–78. doi:[10.1016/j.gde.2011.12.001](https://doi.org/10.1016/j.gde.2011.12.001)
- Antequera F, Boyes J, Bird A (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 62:503–514
- Augui S, Nora EP, Heard E (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* 12:429–442. doi:[10.1038/nrg2987](https://doi.org/10.1038/nrg2987)

- Berger AH, Pandolfi PP (2011) Haplo-insufficiency: a driving force in cancer. *J Pathol* 223:137–146. doi:[10.1002/path.2800](https://doi.org/10.1002/path.2800)
- Berletch JB, Yang F, Distèche CM (2010) Escape from X inactivation in mice and humans. *Genome Biol* 11:213. doi:[10.1186/gb-2010-11-6-213](https://doi.org/10.1186/gb-2010-11-6-213), gb-2010-11-6-213 [pii]
- Bix M, Locksley RM (1998) Independent and epigenetic regulation of the interleukin-4 alleles in CD4⁺ T cells. *Science* 281:1352–1354
- Bock C, Halachev K, Buch J, Lengauer T (2009) EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. *Genome Biol* 10:R14
- Bolduc V, Chagnon P, Provost S, Dube MP, Belisle C, Gingras M, Mollica L, Busque L (2008) No evidence that skewing of X chromosome inactivation patterns is transmitted to offspring in humans. *J Clin Invest* 118:333–341. doi:[10.1172/JCI33166](https://doi.org/10.1172/JCI33166)
- Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434:400–404
- Chadwick LH, Willard HF (2005) Genetic and parent-of-origin influences on X chromosome choice in Xce heterozygous mice. *Mamm Genome* 16:691–699. doi:[10.1007/s00335-005-0059-2](https://doi.org/10.1007/s00335-005-0059-2)
- Chadwick LH, Pertz LM, Broman KW, Bartolomei MS, Willard HF (2006) Genetic control of X chromosome inactivation in mice: definition of the Xce candidate interval. *Genetics* 173:2103–2110. doi:[10.1534/genetics.105.054882](https://doi.org/10.1534/genetics.105.054882)
- Chess A, Simon I, Cedar H, Axel R (1994) Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78:823–834
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. *Nat Genet* 32:432–437
- Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO (2007) A genome-wide association study of global gene expression. *Nat Genet* 39:1202–1207. doi:[10.1038/ng2109](https://doi.org/10.1038/ng2109)
- Farago M, Rosenbluh C, Tevlin M, Fraenkel S, Schlesinger S, Masika H, Gouzman M, Teng G, Schatz D, Rais Y, Hanna JH, Mildner A, Jung S, Mostoslavsky G, Cedar H, Bergman Y (2012) Clonal allelic predetermination of immunoglobulin-kappa rearrangement. *Nature* 490:561. doi:[10.1038/nature11496](https://doi.org/10.1038/nature11496), nature11496 [pii]
- Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet* 24:408–415. doi:[10.1016/j.tig.2008.06.001](https://doi.org/10.1016/j.tig.2008.06.001), S0168-9525(08)00177-7 [pii]
- Gimelbrant AA, Chess A (2006) An epigenetic state associated with areas of gene duplication. *Genome Res* 16:723–729
- Gimelbrant AA, Ensminger AW, Qi P, Zucker J, Chess A (2005) Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. *J Biol Chem* 280:1354–1359
- Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318:1136–1140
- Glaser RL, Ramsay JP, Morison IM (2006) The imprinted gene and parent-of-origin effect database now includes parental origin of de novo mutations. *Nucleic Acids Res* 34:D29–D31. doi:[10.1093/nar/gkj101](https://doi.org/10.1093/nar/gkj101), 34/suppl_1/D29 [pii]
- Gordon GJ, Rockwell GN, Jensen RV, Rheinwald JG, Glickman JN, Aronson JP, Pottorf BJ, Nitz MD, Richards WG, Sugarbaker DJ, Bueno R (2005) Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. *Am J Pathol* 166:1827–1840. doi:[10.1016/S0002-9440\(10\)62492-3](https://doi.org/10.1016/S0002-9440(10)62492-3)
- Hattori D, Millard SS, Wojtowicz WM, Zipursky SL (2008) Dscam-mediated cell recognition regulates neural circuit formation. *Annu Rev Cell Dev Biol* 24:597–620. doi:[10.1146/annurev.cellbio.24.110707.175250](https://doi.org/10.1146/annurev.cellbio.24.110707.175250)
- Heard E, Clerc P, Avner P (1997) X-chromosome inactivation in mammals. *Annu Rev Genet* 31:571–610
- Hollander GA, Zuklys S, Morel C, Mizoguchi E, Mobisson K, Simpson S, Terhorst C, Wishart W, Golan DE, Bhan AK, Burakoff SJ (1998) Monoallelic expression of the interleukin-2 locus. *Science* 279:2118–2121

- Hu B, Castillo E, Harewood L, Ostano P, Reymond A, Dummer R, Raffoul W, Hoetzenecker W, Hofbauer GF, Dotto GP (2012) Multifocal epithelial tumors and field cancerization from loss of mesenchymal CSL signaling. *Cell* 149:1207–1220. doi:[10.1016/j.cell.2012.03.048](https://doi.org/10.1016/j.cell.2012.03.048)
- Hu-Li J, Pannetier C, Guo L, Lohning M, Gu H, Watson C, Assenmacher M, Radbruch A, Paul WE (2001) Regulation of expression of IL-4 alleles: analysis using a chimeric GFP/IL-4 gene. *Immunity* 14:1–11
- Hull J, Campino S, Rowlands K, Chan MS, Copley RR, Taylor MS, Rockett K, Elvidge G, Keating B, Knight J, Kwiatkowski D (2007) Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* 3:e99. doi:[10.1371/journal.pgen.0030099](https://doi.org/10.1371/journal.pgen.0030099)
- Jeffries AR, Perfect LW, Ledderose J, Schalkwyk LC, Bray NJ, Mill J, Price J (2012) Stochastic choice of allelic expression in human neural stem cells. *Stem Cells* 30:1938–1947. doi:[10.1002/stem.1155](https://doi.org/10.1002/stem.1155)
- Jeon Y, Sarma K, Lee JT (2012) New and Xisting regulatory mechanisms of X chromosome inactivation. *Curr Opin Genet Dev* 22:62–71. doi:[10.1016/j.gde.2012.02.007](https://doi.org/10.1016/j.gde.2012.02.007)
- Jung D, Giallourakis C, Mostoslavsky R, Alt FW (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 24:541–570
- Kelsey G, Bartolomei MS (2012) Imprinted genes . . . and the number is? *PLoS Genet* 8:e1002601. doi:[10.1371/journal.pgen.1002601](https://doi.org/10.1371/journal.pgen.1002601)
- Kirchgesner CU, Warren ST, Willard HF (1995) X inactivation of the FMR1 fragile X mental retardation gene. *J Med Genet* 32:925–929
- Ko J, Lee YH, Hwang SY, Lee YS, Shin SM, Hwang JH, Kim J, Kim YW, Jang SW, Ryoo ZY, Kim IK, Namkoong SE, Kim JW (2003) Identification and differential expression of novel human cervical cancer oncogene HCCR-2 in human cancers and its involvement in p53 stabilization. *Oncogene* 22:4679–4689. doi:[10.1038/sj.onc.1206624](https://doi.org/10.1038/sj.onc.1206624)
- Kreso A, O'Brien CA, van Galen P, Gan O, Notta F, Brown AM, Ng K, Ma J, Wienholds E, Dunant C, Pollett A, Gallinger S, McPherson J, Mullighan CG, Shibata D, Dick JE (2012) Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 339:543. doi:[10.1126/science.1227670](https://doi.org/10.1126/science.1227670), [science.1227670](https://doi.org/10.1126/science.1227670) [pii]
- Lamb AN, Rosenfeld JA, Neill NJ, Talkowski ME, Blumenthal I, Girirajan S, Keelean-Fuller D, Fan Z, Pouncey J, Stevens C, Mackay-Loder L, Terespolsky D, Bader PI, Rosenbaum K, Vallee SE, Moeschler JB, Ladda R, Sell S, Martin J, Ryan S, Jones MC, Moran R, Shealy A, Madan-Khetarpal S, McConnell J, Surti U, Delahaye A, Heron-Longe B, Pipiras E, Benzacken B, Passemard S, Verloes A, Isidor B, Le Caignec C, Glew GM, Opheim KE, Descartes M, Eichler EE, Morton CC, Gusella JF, Schultz RA, Ballif BC, Shaffer LG (2012) Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. *Hum Mutat* 33:728–740. doi:[10.1002/humu.22037](https://doi.org/10.1002/humu.22037)
- Lee SH, Harold D, Nyholt DR, Goddard ME, Zondervan KT, Williams J, Montgomery GW, Wray NR, Visscher PM (2013) Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 22:832–841. doi:[10.1093/hmg/dds491](https://doi.org/10.1093/hmg/dds491)
- Li SM, Valo Z, Wang J, Gao H, Bowers CW, Singer-Sam J (2012) Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS One* 7:e31751. doi:[10.1371/journal.pone.0031751](https://doi.org/10.1371/journal.pone.0031751), [PONE-D-11-22867](https://doi.org/10.1371/journal.pone.0031751) [pii]
- Li Q, Seo JH, Stranger B, McKenna A, Pe'er I, Laframboise T, Brown M, Tyekucheva S, Freedman ML (2013) Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 152:633–641. doi:[10.1016/j.cell.2012.12.034](https://doi.org/10.1016/j.cell.2012.12.034)
- Lyons MF (1961) Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190:372–373
- Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C, Barnea G, Kellis M, Lomvardas S (2011) An epigenetic signature for monoallelic olfactory receptor expression. *Cell* 145:555–570. doi:[10.1016/j.cell.2011.03.040](https://doi.org/10.1016/j.cell.2011.03.040), [S0092-8674\(11\)00374-6](https://doi.org/10.1016/j.cell.2011.03.040) [pii]
- Meng X, Lu X, Morris CA, Keating MT (1998) A novel human gene FKBP6 is deleted in Williams syndrome. *Genomics* 52:130–137. doi:[10.1006/geno.1998.5412](https://doi.org/10.1006/geno.1998.5412)

- Miyanari Y, Torres-Padilla ME (2012) Control of ground-state pluripotency by allelic regulation of Nanog. *Nature* 483:470–473. doi:[10.1038/nature10807](https://doi.org/10.1038/nature10807)
- Necsulea A, Semon M, Duret L, Hurst LD (2009) Monoallelic expression and tissue specificity are associated with high crossover rates. *Trends Genet* 25:519–522
- Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, Lavergne K, Villeneuve A, Gaudin T, Brandstrom H, Beck A, Verner A, Kingsley J, Harmsen E, Labuda D, Morgan K, Vohl MC, Naumova AK, Sinnett D, Hudson TJ (2004) A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 16:184–193
- Pereira JP, Girard R, Chaby R, Cumano A, Vieira P (2003) Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nat Immunol* 4:464–470
- Pernis B, Chiappino G, Kelus AS, Gell PG (1965) Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J Exp Med* 122:853–876
- Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK (2010) Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464:768–772. doi:[10.1038/nature08872](https://doi.org/10.1038/nature08872)
- Plagnol V, Uz E, Wallace C, Stevens H, Clayton D, Ozelik T, Todd JA (2008) Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses. *PLoS One* 3:e2966. doi:[10.1371/journal.pone.0002966](https://doi.org/10.1371/journal.pone.0002966)
- Plenge RM, Percec I, Nadeau JH, Willard HF (2000) Expression-based assay of an X-linked gene to examine effects of the X-controlling element (Xce) locus. *Mamm Genome* 11:405–408
- Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF (2002) Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am J Hum Genet* 71:168–173. doi:[10.1086/341123](https://doi.org/10.1086/341123)
- Ratushny V, Gober MD, Hick R, Ridky TW, Seykora JT (2012) From keratinocyte to cancer: the pathogenesis and modeling of cutaneous squamous cell carcinoma. *J Clin Invest* 122:464–472. doi:[10.1172/JCI57415](https://doi.org/10.1172/JCI57415)
- Saferali A, Grundberg E, Berlivet S, Beauchemin H, Morcos L, Polychronakos C, Pastinen T, Graham J, McNeeny B, Naumova AK (2010) Cell culture-induced aberrant methylation of the imprinted IG DMR in human lymphoblastoid cell lines. *Epigenetics* 5:50–60
- Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS (2010) A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10:59–64. doi:[10.1038/nrc2771](https://doi.org/10.1038/nrc2771)
- Schlesinger S, Selig S, Bergman Y, Cedar H (2009) Allelic inactivation of rDNA loci. *Genes Dev* 23:2437–2447
- Serizawa S, Ishii T, Nakatani H, Tsuboi A, Nagawa F, Asano M, Sudo K, Sakagami J, Sakano H, Ijiri T, Matsuda Y, Suzuki M, Yamamori T, Iwakura Y (2000) Mutually exclusive expression of odorant receptor transgenes. *Nat Neurosci* 3:687–693
- Slaughter DP, Southwick HW, Smejkal W (1953) Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 6:963–968
- Takizawa T, Gudla PR, Guo L, Lockett S, Misteli T (2008) Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. *Genes Dev* 22:489–498
- Tonon L, Bergamaschi G, Dellavecchia C, Rosti V, Lucotti C, Malabarba L, Novella A, Vercesi E, Frassoni F, Cazzola M (1998) Unbalanced X-chromosome inactivation in haemopoietic cells from normal women. *Br J Haematol* 102:996–1003
- Wang X, Soloway PD, Clark AG (2010) Paternally biased X inactivation in mouse neonatal brain. *Genome Biol* 11:R79. doi:[10.1186/gb-2010-11-7-r79](https://doi.org/10.1186/gb-2010-11-7-r79)
- Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW (2002) Allelic variation in human gene expression. *Science* 297:1143
- Yang F, Babak T, Shendure J, Distech CM (2010) Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* 20:614–622. doi:[10.1101/gr.103200.109](https://doi.org/10.1101/gr.103200.109), gr.103200.109 [pii]
- Zhang Y, Rohde C, Reinhardt R, Voelcker-Rehage C, Jeltsch A (2009) Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol* 10:R138. doi:[10.1186/gb-2009-10-12-r138](https://doi.org/10.1186/gb-2009-10-12-r138)

- Zuo T, Wang L, Morrison C, Chang X, Zhang H, Li W, Liu Y, Wang Y, Liu X, Chan MW, Liu JQ, Love R, Liu CG, Godfrey V, Shen R, Huang TH, Yang T, Park BK, Wang CY, Zheng P, Liu Y (2007) FOXP3 is an X-linked breast cancer suppressor gene and an important repressor of the HER-2/ErbB2 oncogene. *Cell* 129:1275–1286
- Zwemer LM, Zak A, Thompson BR, Kirby A, Daly MJ, Chess A, Gimelbrant AA (2012) Autosomal monoallelic expression in the mouse. *Genome Biol* 13:R10. doi:[10.1186/gb-2012-13-2-r10](https://doi.org/10.1186/gb-2012-13-2-r10), gb-2012-13-2-r10 [pii]

Part II
Epigenetic Variation in Health
and Disease

Chapter 7

Recurrent CNVs in the Etiology of Epigenetic Neurodevelopmental Disorders

Janine M. LaSalle and Mohammad Saharul Islam

Abstract Recurrent copy number variations (CNVs) are structural large gains and losses of chromosomal domains that are emerging as a major contributor to human neurodevelopmental disorders, such as intellectual disability, autism, schizophrenia, and bipolar disorders. Among the most commonly causal CNVs observed in neurodevelopmental disorders are the rearrangements of proximal chromosome 15q, resulting in Angelman, Prader–Willi, and 15q duplication syndromes. This locus also involves multiple epigenetic layers that influence parental imprinting in the inheritance of these disorders. This chapter summarizes the known CNVs associated with human neurodevelopmental disorders and discusses how epigenetic mechanisms play a role in regulating the genes implicated in these loci. Furthermore, we discuss the epigenetic layer of DNA methylation and its potential role in breakpoint instability in recent primate evolution.

7.1 Introduction

Chromosomal rearrangements include several different classes of large genetic changes, such as deletions, duplications, inversions, and translocations. Chromosomal rearrangements are created when broken DNA double helices at two different locations in the genome are joined in a repair attempt. This results in chromosomal gains or losses, novel structural variations, or a different gene order on the chromosome (Griffiths et al. 1999). Structural variation of the human genome has received recent attention because of the ability to sequence across new chromosomal rearrangements using paired-end next-generation sequencing (Feuk et al. 2006; Freeman et al. 2006). Chromosome rearrangements have been

J.M. LaSalle (✉) • M.S. Islam

Medical Microbiology and Immunology, Genome Center, MIND Institute, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

e-mail: jmlasalle@ucdavis.edu

implicated in Mendelian disease inheritance, but also complex diseases and benign variants with unknown phenotypic consequences (Lupski 2006; Lupski and Stankiewicz 2005). Interestingly, many recurrent and nonrecurrent genomic rearrangements result in copy number variation (CNV) of critical genes involved in the pathogenesis of neurodevelopmental disorders. The genetic basis of disorders involving neurocognitive and behavioral phenotypes such as autism and schizophrenia shows remarkable overlaps in some of the recurrent CNVs. Many of the regions implicated in neurocognitive behavior also involve epigenetic regulation, such as imprinted or X chromosome inactivation, or involve genes involved in epigenetic pathways.

In this chapter, we summarize the CNVs and implicated genes in common and complex neurological traits, with a particular focus on the 15q disorders. In addition, we discuss how epigenetic mechanisms are involved in both the etiology of the chromosomal rearrangements as well as their variable manifestation in disease phenotypes. Lastly, we focus on the evolutionary forces that appear to be acting on a particular hot spot of genomic rearrangements, the 15q11–q13 locus implicated in multiple human neurodevelopmental disorders.

7.1.1 CNVs

Genetic rearrangements that alter chromosomal structure, including inversions and some translocations, may also result in gene copy number differences (Hurles et al. 2008). Duplications, deletions, and complex multisite variants (Feuk et al. 2006), collectively referred to as CNVs, are frequently found in humans and other mammals (Feuk et al. 2006; Freeman et al. 2006). CNV is the most prevalent type of structural variation in the human genome (Redon et al. 2006). As much as 12 % of the human genome and thousands of genes are variable in copy number, and this diversity is likely to be responsible for a significant proportion of normal phenotypic variation (Carter 2007). Annotated CNVs are mostly larger (typically >50 kb) and intermediate-size structural variation (>500 bp) (Carter 2007; Pinto et al. 2007).

CNVs can be *de novo* or familial, with the former more likely to contribute to the development of sporadic genomic disorders (McCarroll et al. 2008). Duplication or deletion of large chromosomal segments can disrupt a variable number of genes, resulting in alternate gene products or changes in allelic expression. Moreover, the disruption of distal regulatory regions in the genome created by CNVs or structural rearrangements can also lead to altered gene expression. Duplications and deletions of genes affecting inflammatory response, immunity, olfactory function, and cell proliferation might have been fixed by positive selection and involved in the adaptive phenotypic differentiation of humans, mice, and chimpanzees (Nguyen et al. 2006; Perry et al. 2008; Schaschl et al. 2009; Young et al. 2008). This diversity is likely to be responsible for a significant proportion of phenotypic variation in humans (Carter 2007).

7.1.2 *Neurodevelopmental Disorders*

Neurodevelopmental disorders (NDs) involve the abnormal growth and development of the brain or central nervous system, which ultimately affects cognition and/or behavior throughout life. Neurodevelopmental disorders are characterized by neurological and psychiatric signs seen in the course of brain development—from conception to early adulthood (Grayton et al. 2012). These disorders encompass a variety of signs and symptoms including a range of cognitive impairments resulting in intellectual disabilities (ID) that are often first characterized as developmental delay (DD). In addition, NDs also include a wide range of associated behavioral abnormalities (such as hyperphagia or aggression), psychotic symptoms, autism spectrum disorder (ASD), sensory impairment, seizures, motor dysfunction, and speech and language difficulties (Lee and Lupski 2006).

The chromosomal rearrangements observed in recurrent CNVs are a source of interindividual genetic variation that could explain variable penetrance of inherited (Mendelian and polygenic) diseases and phenotypic expression of sporadic traits (Beckmann et al. 2007). Table 7.1 summarizes specific CNVs that have been identified and are suspected to be the genomic cause of several neurodevelopmental disorders, including autism, intellectual disability, epilepsy, attention deficit hyperactivity disorder (ADHD), and schizophrenia (Helbig et al. 2009; Mefford et al. 2008; Sebat et al. 2007; Stefansson et al. 2008; Walsh et al. 2008). However, the genetic inheritance is complicated by both variable expressivity and incomplete penetrance associated with specific CNVs. Variable expressivity means that the same CNV may have different clinical manifestations, such as resulting in autism or epilepsy or schizophrenia in different individuals. Incomplete penetrance refers to the inheritance patterns of CNVs that are found in affected individuals but also at a lower frequency in apparently unaffected controls. Since the genetic correlation between specific CNVs and specific neurodevelopmental disorders is not absolute, epigenetic mechanisms are important to consider in their etiology.

7.1.3 *Epigenetic Mechanisms Influencing CNVs*

Epigenetics can be defined as modifications to nucleotides or chromosomes that do not change the genetic sequence, but can affect gene expression and phenotypic outcome by a variety of mechanisms. DNA methylation refers to addition of methyl groups to CpG dinucleotides in the mammalian genome. DNA methylation within the promoter of a gene is almost always associated with repression, while methylation levels in gene bodies and non-genic regions are actually positively associated with expression (Lister et al. 2009; Rauch et al. 2009). In addition to DNA modification, the histone core proteins H3 and H4 are both heavily posttranslationally modified on their N-terminal tails resulting in an epigenetic “histone code” (Margueron et al. 2005; Wang et al. 2004b). Histone marks of H3K4me3 and

Table 7.1 Recurrent CNV associated with neurodevelopmental disorders

Locus	CNV type	Syndromic or variable	Syndrome or symptoms	Implicated genes	Epigenetic involvement	References
1p36	Del	Syndromic	ID, epilepsy, growth delay, hearing, and/or vision loss	<i>GNB1, CALML6, TMEM52, C10RF222, KIAA1751</i>	SKI implicated in chromatin remodeling	Shapira et al. 1997; Gajeccka et al. 2010; Rosenfeld et al. 2010
			PWS-like obesity and hyperphagia	<i>GABRD, PRKCZ, SKI</i>		Baranek and Alanasoski 2012; D'Angelo et al. 2006; D'Angelo et al. 2009; D'Angelo et al. 2010; Rosenfeld et al. 2010
1q21.1	Del	Variably expressive	Schizophrenia, ID, ASD, ADHD, epilepsy	<i>GJA8, GJA5, BCL9, ACP6, GPR89B, GPR89C, PRKAB2, FMO5, CHD1L</i>	Suspected but unknown	ISC 2008; Brunetti-Pierri et al. 2008; Brzustowicz et al. 2000; De Vries et al. 2005; Gurling et al. 2001; Hwu et al. 2003; Ni et al. 2007; Sharp et al. 2006; Stefansson et al. 2008; Walsh et al. 2008; Kim et al. 2008; Zheng et al. 2006
2p16.3	Del	Variably expressive	ASD, ID, schizophrenia	<i>NRXN1</i>	Unknown	Feng et al. 2006; Kim et al. 2008; Kirov et al. 2008; Rujescu et al. 2009; Walsh et al. 2008; Zahir et al. 2008
3q29	Del	Variably expressive	ID, ASD, schizophrenia, bipolar disorder	<i>PAK2, DLG1, TFRG, BDH1, SENPS, NCBP2, MFI2</i>	Unknown	Mulle et al. 2010; Willatt et al. 2005
6q22.1	Del	Variably expressive	ID, microcephaly	<i>MARCKS, HDAC2, H3S3T5</i>	HDAC2 is histone deacetylase	Rosenfeld et al. 2012; Chen et al. 2006
7q31	Del	Syndromic	ASD, language disorder	<i>SPCH1, FOXP2</i>	Tissue-specific methylation of <i>FOXP2</i>	IMGSAC 2001; Ashley-Koch et al. 1999; Consortium 1998; Fisher et al. 1998; Lai et al. 2000; Rice et al. 2012; Tolosa et al. 2010
7q11.23	Del	Variably expressive	Williams–Beuren syndrome	<i>CLIP2, ELN, GTF2I, GTF2IRD1, LIMK1</i>	WBSCR22, MTFase WBSCR9, chromatin remodeling	Curran et al. 1993; Johnson et al. 1976; Pober 2010; Peoples et al. 1998; Doll and Grzeschik 2001

16p11.2	Dup/Del	Variably expressive	Autism, schizophrenia, ADHD	<i>KCTD13</i>	Unknown	McCarthy et al. 2009; Shinawi et al. 2010; Kim et al. 2008; Golzio et al. 2012
17p13.3	Del	Syndromic	Miller–Dieker syn- drome and lissencephaly	<i>PAFAH1B1 (LIS1)</i>	Unknown	Cardoso et al. 2003; Dobyns et al. 1993
17p11.2	Del	Syndromic	Smith–Magenis syndrome	<i>RAI1</i>	Unknown	Chen et al. 1997; Girirajan et al. 2006; Greenberg et al. 1991; Greenberg et al. 1996; Shaw et al. 2004
17q11.2	Del	Syndromic	Neurofibromatosis type 1	<i>NF1</i>	NF1 is epigenetic TF implicated in brain tumors	Chossien et al. 1997; Forbes et al. 2004; Korf 2002; López-Correa et al. 2001; Rasmussen et al. 1998; Nischal et al. 2012
22q11.2	Del	Variably expressive	ID	<i>COMT, TBX1, GNB1L, ZDDHC8, DGCR8, RANBP1, and CDC45L</i>	<i>DGCR8, miR-185 tar- get DNMT1</i>	Bearden et al. 2001; El Tahir et al. 2004; Lynch et al. 1995; Meechan et al. 2009; Moss et al. 1999; Gurling et al. 2001; Paylor et al. 2006; Williams et al. 2008; Zhang et al. 2011; Earls et al. 2012
Xq28	Dup	Syndromic	ID, autism, immune deficiencies	<i>MECP2</i>	XCI, MeCP2, methyl binding	del Gaudio et al. 2006; Meins et al. 2005; Van Esch et al. 2005a
15q11–q13	Mat dup	Variably expressive	Autism, ID, epilepsy	<i>UBE3A</i>	<i>UBE3A</i> imprinting	Battaglia 2008; Bunday et al. 1994; Cook et al. 1997; Kirov et al. 2008; Moeschler et al. 2002
15q11–q13	Mat del	Syndromic	Angelman syndrome	<i>UBE3A</i>	<i>UBE3A</i> imprinting	Andersen et al. 2001; Cassidy et al. 2000; Fang et al. 1999; Huibregtse et al. 1995; Kishino et al. 1997; Knoll et al. 1989; Lalande and Calciano 2007a; Lossie et al. 2001; Malcolm et al. 1991; Matsuura et al. 1997; Penner et al. 1993; Yamasaki et al. 2003b

(continued)

Table 7.1 (continued)

Locus	CNV type	Syndromic or variable	Syndrome or symptoms	Implicated genes	Epigenetic involvement	References
15q11-q13	Pat del	Syndromic	Prader-Willi syndrome	<i>HBI185</i> and <i>IncRNA host</i>	Imprinting of large neuronal transcript from PWS-IC	Cassidy and Driscoll 2009; Chamberlain and Lalande 2010; De Smith et al. 2009; Doe et al. 2009; Kishore and Stamm 2006; Hogart et al. 2009; Brunetti-Pierri et al. 2008; Vitali et al. 2010
15q11-q13	Mat trip	Variably expressive	Idic(15)	<i>UBE3A</i> , <i>GABRB3</i> , <i>SNRPN</i>	<i>UBE3A</i> imprinting, long-range <i>cis</i> and <i>trans</i> effects	Battaglia 2005; Chamberlain and Lalande 2010; Cook et al. 1998; Hogart et al. 2009; Hogart et al. 2010; Meguro-Horike et al. 2011b; Repetto et al. 1998; Robinson et al. 1998; Schroer et al. 1998; Wandstrat et al. 1998

H3K36me3 and acetylation of H3 and H4 at multiple sites are observed at active genes, while H3K27me3 and H3K9me3 are markers of a repressed chromatin state (Wang et al. 2004b). In addition, multiple histone modifications and variants influence the repair of double-stranded breaks, so the histone state can influence susceptibility of the genome to rearrangements (Downs et al. 2007). In yeast, heterochromatin is less accessible to DNA repair events that can lead to chromosomal rearrangements. Furthermore, genomic regions of differential low methylation and silent gene expression were highly enriched for breakpoints and CNVs (Tang et al. 2012), and the regions of the genome with hypomethylation show increases in structural rearrangements (Li et al. 2012). Large-scale domains of partial methylation called PMDs have been observed in cancer and primary embryonic cell lines and are regions of heterochromatic silent marks and low levels of transcription (Li et al. 2012; Lister et al. 2009; Schroeder et al. 2011). Together, the emerging evidence suggests that the local chromatin environment can influence the occurrence of structural rearrangements selectively in heterochromatin.

Two classic examples of epigenetic mechanisms influencing phenotype are parental imprinting and X chromosome inactivation. Autosomal regions affected by parental imprinting exhibit allele-specific differences in gene expression based on parental origin (Reik and Walter 2001; Morison et al. 2005). Opposite patterns of DNA methylation and histone states differentially mark parentally imprinted regions primarily through one or more imprinting control regions (ICR) that determine the parental-specificity of gene expression in these loci (Soejima and Wagstaff 2005). Similarly, the X chromosome in females is subject to X chromosome inactivation (XCI) in order to achieve dosage compensation with males (Payer and Lee 2008). A detailed review of XCI may be found in Chap. 2. The main difference in XCI from parental imprinting is that after implantation the choice of X chromosome is random, resulting in an expected 50 % maternally and 50 % paternally expressed alleles. However, XCI can be skewed in carriers of X-linked disease genes or CNVs (Robinson et al. 2001; Knudsen et al. 2006). Even in women who are not obvious carriers for disease-causing variants, skewed XCI can be observed, likely because of stochastic or benign genetic variants (Hatakeyama et al. 2004). Therefore, CNVs occurring on imprinted loci or the X chromosome may have more severe effects from the chromosomal gain or loss because only one allele is active.

DNA methylation is an important epigenetic mechanism not only in the phenotypic manifestation of CNVs but also potentially in their etiology. The human genome is the most highly methylated and highly repetitive of mammalian genomes, and the high levels of DNA methylation have been considered to be a genome defense mechanism against the spread of retrotransposons and low copy repeats (LCRs) (Bestor and Tycko 1996). Importantly, global DNA hypomethylation is associated with genome instability, as observed in many tumor types exhibiting global DNA hypomethylation and in triplet repeat expansion (LaSalle 2011). Furthermore, the 1 % of the human genome characterized as “methylation deserts” was enriched for susceptibility to CNVs (Li et al. 2012). Interestingly, many different environmental exposures such as heavy metals, air pollutants, and persistent organic

pollutants are not known mutagens, but are associated with global DNA hypomethylation and could be indirectly predisposing to CNVs (Baccarelli and Bollati 2009). In support of the possibility, high levels of the organic polychlorinated biphenol PCB-95 and reduced DNA methylation were observed in brain samples from individuals with maternally derived duplication 15q syndrome (Mitchell et al. 2012b).

Additional epigenetic mechanisms with potential relevance to the etiology of CNVs in the human genome are the structural layers of chromatin loops and three-dimensional nuclear organization. The linear chromosomal maps do not necessarily reflect the spatial organization of genes within the interphase nucleus (Kosak and Groudine 2004). The recent ENCODE project has demonstrated that long-range interactions between distant genes on the same or the different chromosomes are quite common in the human genome and are cell type dependent (Dunham et al. 2012; Horike et al. 2005; Meguro-Horike et al. 2011a; Yasui et al. 2011). The structural CCTC-binding factor (CTCF) and the methyl cytosine-binding protein 2 (MeCP2) are both nuclear factors with genome-wide distribution in neurons that regulate chromatin looping and nuclear organization (Kernohan et al. 2010). Therefore, existing CNVs may alter gene expression beyond the breakpoint boundaries because of long-range chromatin effects. In addition, some genetic loci distally located in the genome may be more susceptible to the formation of CNVs because of their colocalization in the interphase nucleus.

7.2 CNVs Associated with Neurodevelopmental Disorders

7.2.1 *CNVs in Chromosome 1 and Associated ND*

Deletions of 1p36 represent the most common syndrome associated with terminal deletions in humans, occurring in 1 in 5,000 live births (Gajecka et al. 2010; Rosenfeld et al. 2010). Different deletion sizes are associated with somewhat different clinical manifestations. Features in common to all 1p36 monosomies include ID, DD, hypotonia, dysmorphic facial features, and microcephaly. Individuals with more proximal 1p36.33 deletions exhibit a phenotype similar to Prader–Willi syndrome (PWS) with hyperphagia and obesity (D’Angelo et al. 2006, 2009, 2010). This proximal locus includes the gene encoding *v-ski* sarcoma viral oncogene homolog (SKI), a transcriptional regulator required for the maintenance of the neural stem cell pool and the development of the *corpus callosum*. SKI appears to function as a transcriptional repressor by interactions with the chromatin-remodeling factor SATB homeobox 2 (SATB2) and histone deacetylase (Baranek and Atanasoski 2012).

Deletions at 1q21.1 show nominal association with schizophrenia (Stefansson et al. 2008). Previously reported 1q21.1 deletions in two cases of ID (De Vries

et al. 2005; Sharp et al. 2006), two autistic individuals (Weiss et al. 2008), and one schizophrenia case (Walsh et al. 2008) are consistent with the shorter form of the deletion identifying the causative locus. In at least four reports (Brzustowicz et al. 2000; Gurling et al. 2001; Hwu et al. 2003; Zheng et al. 2006) the 1q21 locus has been linked to schizophrenia. The 1.35 Mb deleted segment common to both the large and the small form of the 1q21.1 deletion is gene rich (Stefansson et al. 2008), containing 27 known genes, most of which are expressed in the brain (ISC 2008), and previous reports have shown linkage of this locus to ID (Brunetti-Pierri et al. 2008). The gene encoding gap junction protein, alpha 8 (*GJA8*), is expressed in brain and located in a repeat region within the boundary of the 1.35 Mb deletion segment and previously reported as associated with schizophrenia (Ni et al. 2007). Recurrent reciprocal microdeletions and microduplications within 1q21.1 represent novel genomic disorders consisting of microcephaly or macrocephaly, respectively, and can manifest with a range of developmental delay, neuropsychiatric abnormalities, dysmorphic features, and a variety of other congenital anomalies (Brunetti-Pierri et al. 2008).

7.2.2 CNVs in Chromosome 2 and Associated ND

A case study reported that a child with developmental delay, unusual autistic-like behaviors, multiple vertebral anomalies, and an unusual facial appearance was found to have a de novo 321 kb deletion of chromosome 2p16.3 by array comparative genomic hybridization (aCGH) that deleted a 5' portion of *NRXN1*, encoding neurexin 1, a neuronal cell adhesion molecule (Zahir et al. 2008). Another report implicated *NRXN1* in two independent subjects with ASD and a balanced chromosomal rearrangement at 2p16.3 (Kim et al. 2008). Furthermore, *NRXN1* missense variants were associated with autism in a case-control study (Feng et al. 2006). Deletions of 2p16.3 that disrupts *NRXN1* were also observed in schizophrenia (Kirov et al. 2008; Rujescu et al. 2009). A 115-kb deletion on chromosome 2p16.3 disrupting *NRXN1* was also found in identical twins concordant for childhood-onset schizophrenia (Walsh et al. 2008).

7.2.3 CNVs in Chromosome 3q and Associated ND

Patients with autism and 3q29 microdeletion also exhibited ataxia, chest-wall deformity, and long and tapering fingers (Willatt et al. 2005). The 1.5 Mb 3q29 microdeletion encompasses 22 genes, including *PAK2* and *DLG1*, which are autosomal homologs of two known X-linked mental retardation genes, *PAK3* and *DLG3*. Another study found six 3q29 microdeletions among 7,545 schizophrenic subjects compared to 1 out of 39,748 controls, resulting in a statistically significant association with schizophrenia (Mulle et al. 2010).

7.2.4 CNVs in Chromosome 6 and Associated ND

A large 24 Mb deletion in 6q22.1–q23.2 was reported in an infant with DD, microcephaly, facial dysmorphism, pulmonary atresia, and ventricular septal defect (Rosenfeld et al. 2012). More recently, seven individuals with ND refined the deletion locus at 6q22.1 to a 250 kb cluster of four genes, including *MARCKS*, *HDAC2*, and *HS3ST5* which are implicated in neurodevelopment (Rosenfeld et al. 2012). *HDAC2* encodes for histone deacetylase 2, a protein important in epigenetic gene repression.

7.2.5 CNVs in Chromosome 7 and Associated ND

Multiple studies have shown that q31 region on chromosome 7 co-segregates with speech and language disorders. Chromosomal rearrangements involving 7q31 (including translocations, inversions, and a duplication) have been observed in patients with autism (IMGSAC 2001; Ashley-Koch et al. 1999). *SPCH1* on human 7q31 was associated with speech and language disorder (Lai et al. 2000). *SPCH1* interval overlaps with an ~40-cM region identified in a genome screen for susceptibility to autism, a disorder which is often associated with speech and language abnormalities (Consortium 1998; Fisher et al. 1998). The gene encoding the FOXP2 transcription factor is located in the same locus and also implicated in the speech and language disorders. A minimal deletion in mother and son with *FOXP2* haploinsufficiency due to a 1.57-Mb deletion on chromosome 7q31 (Rice et al. 2012) also included genes, *MDFIC* and *PPP1R3A*. The boy had severe childhood apraxia of speech, with poor expressive speech, severely delayed speech acquisition, and inability to laugh, sneeze, or cough spontaneously. He showed mildly impaired cognition, which may have been due to the speech limitations. His 24-year-old mother was similarly, if slightly less, affected. She had a similar early developmental history, with speech apraxia and mild developmental delay. Neither patient had autistic features. Tissue-specific methylation differences have been observed at FOXP2 that may complicate gene association studies (Tolosa et al. 2010).

Several genes encoded in chromosome 7q11.23 are dosage sensitive and play a role in human language. Deletion of chromosome 7q11.23 causes Williams–Beuren syndrome (WBS), characterized by spatial learning deficits and aberrant social behaviors (Bozhenok et al. 2002; Amir et al. 1999). A recurrent microdeletion of around 1.6 Mb results in loss of 28 genes central to WBS, including *CLIP2*, *ELN*, *GTF2I*, *GTF2IRD1*, and *LIMK1* (Pober 2010). Loss of the *ELN* gene, which codes for the protein elastin, is associated with the connective-tissue abnormalities and cardiovascular disease (specifically supralvalvular aortic stenosis and supralvalvular pulmonary stenosis) found in many people with this syndrome (Curran et al. 1993; Johnson et al. 1976). At least two of the relatively uncharacterized genes within the

WBS deletion have predicted epigenetic functions, including WBSCR22 that encodes a putative methyltransferase and WBSCR9 that encodes a transcriptional regulator (Doll and Grzeschik 2001; Merla et al. 2002; Peoples et al. 1998).

7.2.6 CNVs in Chromosome 16 and Associated ND

Recurrent rearrangements of the 16p11.2 locus have been observed exhibiting both reduced penetrance and variable expressivity. Both microdeletions and microduplications of chromosome 16p11.2 have been implicated in autism (Weiss et al. 2008), from a study within the Autism Genetic Resource Exchange (AGRE) population that identified five instances of de novo deletions on chromosome 16p11.2 as well as reciprocal duplication of the 593 kb deleted region. However, these microdeletions and duplications are also observed in around 1 % of controls (IMGSAC 2001; de Kovel et al. 2010; McCarthy et al. 2009; Shinawi et al. 2010; Weiss et al. 2008). This deleted or duplicated region on 16p11.2 contains 25 annotated genes. Recently some reports also showed that the microduplications of 16p11.2 are associated with schizophrenia and ADHD (McCarthy et al. 2009; Shinawi et al. 2010). The gene *KCTD13*, encoding polymerase delta-interacting protein 1 and predicted to be involved in neurogenesis, was identified as the sole contributor to the microcephaly phenotype of 16q11.2 CNVs in a screen of overexpression in zebrafish embryos (Golzio et al. 2012).

7.2.7 CNVs in Chromosome 17 and Associated ND

Microdeletion of 17p13.3 causes Miller–Dieker lissencephaly syndrome (MDLS; OMIM # 247200). The MDLS deleted region includes the lissencephaly-1 (*LIS1*) gene, *PAFAH1B1* (Dobyns et al. 1993). MDLS is characterized by the brain malformation associated with deletion of *LIS1* (*PAFAH1B1*), and includes abnormal facial appearance and severe ID. Usually MDLS patients have large deletion intervals (more than 1.3 Mb), which show a more severe grade of lissencephaly, likely due to the inclusion of particular genes other than *PAFAH1B1* in the deletion interval (Cardoso et al. 2003).

Microdeletions of 3.7 Mb on 17p11.2 cause Smith–Magenis syndrome (SMS; OMIM #182290), a developmental disorder (Chen et al. 1997; Greenberg et al. 1991; Shaw et al. 2004). SMS is characterized by multiple congenital anomalies, mental retardation, a variable degree of developmental delay, behavioral and physical abnormalities such as hearing impairment, and minor skeletal and craniofacial defects (Greenberg et al. 1991, 1996). Although this region contains multiple genes, the loss of the retinoic acid induced 1 or *RAI1* is responsible for most of the characteristic features of this condition (Girirajan et al. 2006). Also, other genes within the chromosome 17 contribute to the variability and

severity of the clinical features. The loss of other genes in the deleted region may help explain why the features of SMS vary among affected individuals.

CNVs are also observed in the q11.2 region on chromosome 17. 1.5 Mb recurrent interstitial microdeletions, which include the NF1 tumor-suppressor gene, are found in 5–20 % of patients with autosomal dominant neurofibromatosis type 1 (NF1; OMIM #162200) (Cnossen et al. 1997; Rasmussen et al. 1998). This recurrent contiguous gene deletion (type I) encompasses at least 13 genes other than NF1, which occurs with rearrangement hot spots contained within the *NFIREP-P1* and *NFIREP-M* LCRs (Forbes et al. 2004; López-Correa et al. 2001). NF1 is primarily characterized by multiple benign nerve-sheath tumors or neurofibromas and pigmentary changes (Korf 2002). NF1 encodes a transcription factor that binds to epigenetically dysregulated targets in human cancers (Nischal et al. 2012).

7.2.8 CNVs in Chromosome 22 and Associated ND

A microdeletion of 22q11.2 causes the velo-cardiofacial syndrome (VCFS; OMIM #192430), which is also known as DiGeorge syndrome (DGS; OMIM #188400). VCFS is associated with developmental delay and a wide range of cognitive and neurological deficits, which include speech, language, memory, and attention (Bearden et al. 2001; El Tahir et al. 2004; Lynch et al. 1995; Moss et al. 1999). The 3 Mb deletion region causing VCFS contains the gene encoding catechol-O-methyl transferase (COMT), an enzyme responsible for the degradation of dopamine. There are many other genes thought to be involved in brain function and neurodevelopment, including *TBX1*, *GNB1L*, *ZDDHC8*, *DGCR8*, *RANBP1*, and *CDC45L* (Meechan et al. 2009; Paylor et al. 2006; Williams et al. 2008). Some VCFS individuals with 22q11.2 deletion were also found to have schizophrenia (Murphy and Owen 2001), suggesting that there could be a link between the other genes of this region and psychiatric disorders. Interestingly, *DGCR8* encodes a miRNA biogenesis protein that regulates miR-185, which in turn targets the DNA methyltransferase gene DNMT1, resulting in global hypomethylation in human glioma (Zhang et al. 2011; Earls et al. 2012).

7.2.9 CNVs in Chromosome X and Associated ND

The X chromosome has both a genetic and epigenetic connection to neurodevelopmental disorders, resulting in differences in disease penetrance in males versus females. Duplications of chromosome Xq28 are observed in males with ID, autism, and immune abnormalities, and the primary gene implicated is *MECP2* (Kirk et al. 2009; Lugtenberg et al. 2009; Prescott et al. 2009; Van Esch et al. 2005b; Velinov et al. 2009). While point mutations in *MECP2* are the most frequent cause of Rett syndrome (RTT) (Amir et al. 1999b), a rare genomic

deletion of the *MECP2* gene was observed (Ravn et al. 2005). RTT (Amir et al. 1999b) is a neurodevelopmental disorder, which mostly affects females. Girls with classical RTT exhibit an apparently normal development of 6–18 months of age, followed by a regressive stage characterized by deceleration of head growth and loss of speech and acquired motor skills. Male cases of RTT are very rare, limited to exceptional cases of somatic mosaicism or X chromosome aneuploidy (Moog et al. 2003). Five critical genes were found within the 200 kb minimal duplication region, including *LICAM* and *MECP2* (del Gaudio et al. 2006; Meins et al. 2005; Van Esch et al. 2005a). The phenotypic severity of Xq28 duplications depended on dosage of *MECP2* more than duplication size. *MECP2* is subject to X chromosome inactivation in females and also encodes an epigenetic factor that binds to methylated DNA.

7.3 15q Chromosomal Rearrangement and Related Neurodevelopmental Disorder

15q11–q13 is a notable chromosome region, because it is both highly epigenetically regulated and characterized by multiple structural abnormalities like deletions, duplications, triplications, and translocations (Hogart et al. 2008). Chromosome 15 is also enriched in segmental LCRs (Bailey et al. 2002) and LCR-mediated misalignment to unequal nonallelic homologous recombination, which generates a series of common breakpoints (BPs) along the 15q11.2–q13 (Christian et al. 1999). There are three BP clusters located in proximal 15q that correspond to complex LCRs within 50–500 kb (Fig. 7.1) (Amos-Landgraf et al. 1999; Makoff and Flomen 2007). The repeats at BP1–BP3 show limited sequence homology to each other, but the two more distal BP clusters (BP4 and BP5) involved a distinct set of LCRs (Makoff and Flomen 2007). This complex structure of tandem and inverted LCRs on proximal 15q contributes to a variety of recurrent and more complex deletions and duplications observed in neurodevelopmental disorders (Hogart et al. 2010). In addition to the complexity of the genetic backbone of chromosome 15q11–q13, the center of this locus is also subject to parental imprinting, an epigenetic phenomenon associated with parental allele-specific differences in gene expression, DNA methylation, and chromatin organization.

The imprinted gene expression observed in 15q11–q13 has important implications for the inheritance patterns of the human diseases associated with this locus. Three different neurodevelopmental disorders map to this region, including PWS, Angelman syndrome (AS), and 15q duplication syndrome (de Kovel et al. 2010; Doornbos et al. 2009; Helbig et al. 2009; Miller et al. 2009; Murthy et al. 2007; Sharp et al. 2008; Stefansson et al. 2008). Large deletions around 5 Mb are responsible for approximately 70 % of cases of PWS and AS. These LCR-mediated deletions occur during meiosis at breakpoints 1–3 (BP1–BP3, Fig. 7.1) (Amos-Landgraf et al. 1999; Christian et al. 1999). PWS results from

15q11–q13 (Knoll et al. 1989), ~10 % by maternal *UBE3A* mutations (Fang et al. 1999), ~10 by paternal uniparental disomy (UPD) of chromosome 15 (Malcolm et al. 1991), and the remainder by microdeletions or epigenetic mutations causing a paternal methylation pattern on the maternally inherited chromosomes, collectively called imprinting mutations (Buiting et al. 2003; Glenn et al. 1993; Reis et al. 1994).

Angelman syndrome patients exhibit ataxia, microcephaly, frequent seizures, and profound learning disabilities coupled with short attention span, absent speech, and characteristic happy demeanor (Cassidy et al. 2000; Lossie et al. 2001). AS language skills are severely impaired, remaining at less than or equivalent of 2 years of age throughout life (Andersen et al. 2001; Penner et al. 1993). The AS critical region is 35 kb telomeric to the PWS critical region (Lalande and Calciano 2007a).

The AS gene, *UBE3A*, encodes a member of a class of functionally related E3 ubiquitin–protein ligases and is expressed from both parental alleles in most tissues (Huibregtse et al. 1995) except the brain, where *UBE3A* is exclusively expressed from the maternally inherited allele (Rougeulle et al. 1998). So, specifically in the postnatal brain, mutation or deletion of maternal *UBE3A* results in a complete loss of the ubiquitin ligase activity (Yamasaki et al. 2003b). The mechanism of *UBE3A* imprinting is complex and involves multiple epigenetic layers. First, a maternally inherited imprinting center (AS-ICR, Fig. 7.1) is required for methylation and silencing of the PWS-ICR on the maternal allele in early development (Perk et al. 2002; Johnstone et al. 2005). Differential histone modifications and binding of MeCP2 are also observed at the PWS-ICR that distinguish maternal and paternal alleles (Thatcher et al. 2005; Fulmer-Smentek and Francke 2001; Gregory et al. 2001; Makedonski et al. 2005). The second requirement for *UBE3A* imprinting is for the neuronal transcription of an extremely long transcript originating from the paternal PWS-ICR through *SNRPN*, multiple noncoding RNA clusters, and ending with an antisense transcript of *UBE3A* (*UBE3A-AS*) (Rougeulle et al. 1998; Yamasaki et al. 2003a; Lalande and Calciano 2007b). While all cell types express the protein coding transcript for *SNRPN*, only postnatal neurons transcribe completely through to the *UBE3A-AS* in order to turn off the paternal allele of *UBE3A*. Interestingly, a recent screen for small molecules that could reestablish paternal *Ube3a* expression in mouse neurons identified several inhibitors of topoisomerases, enzymes that function to change the topology of DNA structures (Huang et al. 2012).

7.4.2 Prader–Willi Syndrome

PWS (OMIM 176270) is primarily a neurodevelopmental disorder, but it also affects multiple organs and metabolism. The opposite of AS, PWS is caused by the loss of paternal 15q11–q13. Most of the PWS patients have a large deletion of paternal 15q11–q13, while other patients have either maternal UPD or an

imprinting defect which lacks the presence of the PWS-ICR (Chamberlain and Lalande 2010). The typical symptoms of PWS are hypotonia and failure to thrive seen in infant, followed by hyperphagia, leading to morbid obesity in childhood. Other common symptoms of PWS include sleep abnormalities, small stature with small hands and feet, and obsessive–compulsive disorder (Cassidy and Driscoll 2009).

While there are multiple paternally expressed protein-coding genes in the PWS locus, including *MKRN3*, *MAGEL2*, *NDN*, and *SNURF-SNRPN*, the most recent evidence from both human and mouse genetics has pointed to locus containing on noncoding repeated units of the C/D box small nucleolar RNA (snoRNA) genes (Hogart et al. 2010). PWS patients lacking the *SNORD116* (HBII-85) snoRNA-encoding locus suffer from the same failure to thrive, hypotonia, and hyperphagia that are observed in patients with larger deletions and maternal UPD (De Smith et al. 2009; Sahoo et al. 2008; Duker et al. 2011). Two different mouse models, both with deletions specifically encompassing *Snord116* not the adjacent *Snord115* cluster, have demonstrated a recognizable phenotype of reduced growth and altered metabolism that characterize the infant-stage PWS phenotype (Skryabin et al. 2007; Ding et al. 2008). While some earlier studies suggested a role for the *SNORD115* cluster in PWS (Doe et al. 2009; Kishore and Stamm 2006), paternal deletion of this locus does not appear sufficient to cause PWS (Ding et al. 2005).

Similar to the epigenetic mechanisms for AS, the epigenetic mechanisms regulating and involving the *Snord116* locus are decidedly complex. *Snord116* transcription is controlled from the PWS-ICR, which is methylated and silenced on the maternal allele in all tissues. In neurons, transcription progresses beyond *Snprn*, through *Snord116* and *Snord115*, eventually reaching the *Ube3a-as*. Interestingly, there is emerging evidence that the different noncoding RNA encoded within *Snord115* and *Snord116* may act in dual roles to open chromatin structure and mediate nucleolar maturation during neural development (Leung et al. 2009; Vitali et al. 2010).

7.4.3 15q Duplication Syndrome

15q duplication syndrome is a clinically identifiable syndrome that occurs in two subtypes based on the type of chromosomal rearrangement. Isodicentric chromosome 15 (idic(15)) refers to a supernumerary chromosome (Battaglia 2005). Isodicentric chromosomes can occur via U-type crossover events in meiosis (Robinson et al. 1998), which form a supernumerary derivative chromosome 15 with two centromeres (Hogart et al. 2010). Interstitial duplications of proximal 15q11.2–13 occur without changing the chromosomal number, and are the reciprocal rearrangement as observed as deletions in PWS and AS, through duplication of BP1–BP3 or BP2–BP3 (Repetto et al. 1998). Some interstitial duplications and idic (15) rearrangements are observed with distal LCR contribution, extending to BP4 and BP5 (Wandstrat et al. 1998).

15q duplications have emerged as the single most common causative CNV found in autism at 1–3 % (Cook et al. 1998; Schroer et al. 1998). In addition to autism, individuals with 15q duplication syndrome exhibit hypotonia at birth, motor skills and language development delay, disabilities in cognitive and learning, and epilepsy (Chamberlain and Lalande 2010). Some individuals also present with anxiety, hyperactivity, and short stature (Battaglia 2005). Individuals with idic (15) are more severe than interstitial duplication patients, because they have three copies of maternal 15q11–q13, whereas interstitial duplication patients have two copies (Battaglia 2008).

One report showed that a boy with autism, epilepsy, ataxia, and an interstitial duplication of 15q had the duplication of the *GABRA5* and *GABRB3* genes (Bundey et al. 1994). Researchers also showed that 15q13.1 duplication spanning *APBA2* in schizophrenia (SZ) patient (Kirov et al. 2008). A common chromosome rearrangement is that an inverted duplication occurs tetrasomy for 15q11–q13, which causes more severe phenotype (Battaglia 2008).

Interestingly, the clinical variability of 15q duplication syndrome is less related to differences in genetic breakpoints than in epigenetic or stochastic factors. Two studies comparing brain expression levels suggest that most 15q-encoded genes do not act according to imprint and copy number, with the exception being *UBE3A* that was increased in all brain samples (Hogart et al. 2009; Scoles et al. 2011). The paternal transcript *SNRPN* and the biallelically expressed *GABRB3* were actually decreased compared to controls, an unexpected result based on predicted imprinting patterns and copy number. A recent study also showed reduced transcript levels of *NDN*, *SNRPN*, *GABRB3*, and *CHRNA7* in 15q duplication neuronal model that recapitulated the altered transcription levels observed in postmortem brain (Meguro-Horike et al. 2011b).

7.5 Evolutionary Considerations of CNVs and Neurodevelopment

Both common and low copy genetic repeats are the major risk factors for recurrent CNVs that arise in ND, as transposons, simple sequence repeats, processed pseudogenes, and tandemly repeated sequences together make up over a third of the human genome (Zepeda-Mendoza et al. 2010). Intriguingly, the primate lineage has quite recently evolved the unique group of LCRs (also called segmental dups) implicated in the recurrent CNV rearrangements. Primate-specific LCRs are unique because their breakpoints are especially CpG dense due to their enrichment in the *Alu* class of repeats (Conrad et al. 2010).

In a comparison of human chromosomal rearrangements between human and chimpanzee (Fig. 7.2), only about half of the human chromosomes show structural divergence from chimpanzee. Chromosome 15q11–q13 emerges as the major hot spot for chromosomal structural differences by this analysis. The divergent

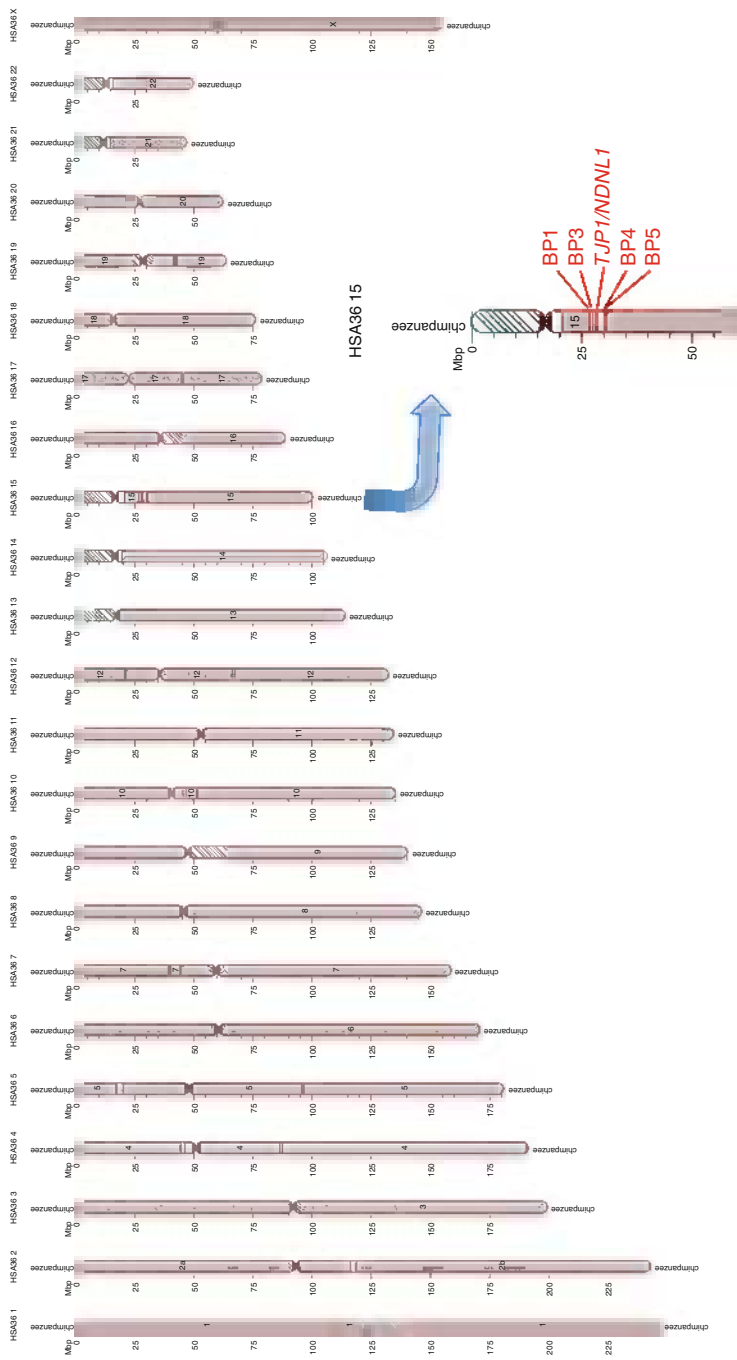


Fig. 7.2 Comparison of human with chimpanzee for structural rearrangements in recent evolution. This image is taken from Evolution Highway (<http://eh-demo.ncsa.uiuc.edu:1715/>). Human chromosome ideograms are shown with *red bars* representing chromosomal structural differences from the chimpanzee genome. While overall there are few changes in chromosomal structural breakpoints between human and chimpanzee, chromosome 15q11–q13 is a hot spot for recent evolutionary changes. The positions of the *red bands* on 15q11–q13 correspond to several breakpoint regions (BP1, 3–5) and gene duplications in *TJP1/NDNL1* (*enlarged inset*)

sequences on 15q11–q13 correspond to the major breakpoint regions discussed in the previous section as being the sites of rearrangement leading to the 15q-associated NDs. The recent evolutionary changes in human proximal 15q may also begin to explain why the various mouse models of 15q-associated NDs do not fully recapitulate the phenotypes, despite a strong conservation of the protein-coding gene sequences and imprinting patterns.

Since 25 % of the total DNA methylation in human cells is at *Alu* sequences (Xie et al. 2009), the epigenetic mechanism of DNA methylation is predicted to be a major factor in predisposing *Alu*-rich LCRs to CNV rearrangements in recent evolution. In support of an epigenetic role in primate LCR origins, the hybrid primate species of white-cheeked gibbons are hypomethylated at *Alu* sites compared to other primates, correlating with increased structural variants with breakpoints corresponding to the lowest methylated *Alu* sites (Carbone et al. 2009).

Therefore, it perhaps is not surprising that LCRs and the CNVs that arise from these *Alu*-rich repeats may play an important role in primate evolution and species differences in levels of brain-expressed transcripts. The transcriptome of human brain shows many brain-specific transcript-level differences compared to other primates (Caceres et al. 2003; Enard et al. 2002). The regions with the biggest differences in gene neighborhood and brain expression between human and chimpanzee (De et al. 2009) are also the human chromosomal regions significantly enriched for LCRs (Jiang et al. 2007), including 15q11–q13.

7.5.1 *Combinations of CNVs and Total CNV Burden in ND*

In addition to their recent evolutionary differences, LCR repeat blocks are also quite polymorphic between individual human genomes. Approximately 3 % of assignable duplications are predicted to be unique to an individual, even though the largest LCR duplication blocks are invariant (Alkan et al. 2009). Since CNVs are found at some level in all human genomes, the assignment of specific CNVs to specific neurodevelopmental disorders is often problematic. More recently, combinations of CNVs have been explored in order to investigate the whole-genome burden of CNVs in individuals with NDs. Whole chromosomal differences may contribute to hypomethylation leading to further rearrangements, so it is important to investigate both environmental contributors to DNA hypomethylation as well as combinations of CNVs in the future.

One recent study examined the total burden of CNV gains and losses in individuals with a range of ND, including autism, ID, or dyslexia, compared to controls and found that the largest CNV burden correlated with the severity of the disorder (Girirajan et al. 2011). These types of studies will be important in the future to improve the understanding of the link between CNVs and ND beyond the syndromic forms. The other question that investigations of total CNV burden raises is what environmental exposures may predispose to CNVs and the DNA hypomethylation associated with increased rearrangements of LCRs. In our recent

analysis of persistent organic pollutant exposures in postmortem brain samples of ND, we observed an unexpected association between PCB-95 exposures and chromosome 15q duplication syndrome and DNA hypomethylation (Mitchell et al. 2012a). Future studies will be needed to determine the causality of specific environmental exposures with increased recurrent CNV occurrences in human cells and in transgenerational studies.

References

- Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41 (10):1061–1067
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl- CpG-binding protein 2. *Nat Genet* 23(2):185–188
- Amos-Landgraf JM, Ji Y, Gottlieb W, Depinet T, Wandstrat AE, Cassidy SB, Driscoll DJ, Rogan PK, Schwartz S, Nicholls RD (1999) Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet* 65(2):370
- Andersen WH, Rasmussen RK, Strømme P (2001) Levels of cognitive and linguistic development in Angelman syndrome: a study of 20 children. *Logoped Phoniatr Vocol* 26(1):2–9
- Ashley-Koch A, Wolpert C, Menold M, Zaeem L, Basu S, Donnelly S, Ravan S, Powell C, Qumsiyeh M, Aylsworth A (1999) Genetic studies of autistic disorder and chromosome 7. *Genomics* 61(3):227–236
- Baccarelli A, Bollati V (2009) Epigenetics and environmental chemicals. *Curr Opin Pediatr* 21 (2):243–251
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. *Science* 297 (5583):1003–1007
- Baranek C, Atanatoski S (2012) Modulating epigenetic mechanisms: the diverse functions of Ski during cortical development. *Epigenetics* 7(7):676–679. doi:10.4161/epi.20590
- Battaglia A (2005) The inv dup (15) or idic (15) syndrome: a clinically recognisable neurogenetic disorder. *Brain Dev* 27(5):365–369
- Battaglia A (2008) The inv dup (15) or idic (15) syndrome (tetrasomy 15q). *Orphanet J Rare Dis* 3:30
- Bearden CE, Woodin MF, Wang PP, Moss E, McDonald-McGinn D, Zackai E, Emmanuel B, Cannon TD (2001) The neurocognitive phenotype of the 22q11. 2 deletion syndrome: selective deficit in visual-spatial memory. *J Clin Exp Neuropsychol* 23(4):447–464
- Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8(8):639–646
- Bestor TH, Tycko B (1996) Creation of genomic methylation patterns. *Nat Genet* 12(4):363–367
- Bozhenok L, Wade PA, Varga-Weisz P (2002) WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J* 21(9):2231–2241
- Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, Lalani SR, Graham B, Lee B, Shinawi M (2008) Recurrent reciprocal 1q21. 1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 40(12):1466–1471

- Brzustowicz LM, Hodgkinson KA, Chow EWC, Honer WG, Bassett AS (2000) Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21–q22. *Science* 288 (5466):678–682
- Buiting K, Gross S, Lich C, Gillissen-Kaesbach G, el-Maarri O, Horsthemke B (2003) Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. *Am J Hum Genet* 72(3):571–577
- Bundey S, Hardy C, Vickers S, Kilpatrick M, Corbett J (1994) Duplication of the 15q11–13 region in a patient with autism, epilepsy and ataxia. *Dev Med Child Neurol* 36(8):736–742
- Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc Natl Acad Sci U S A* 100(22):13030–13035. doi:[10.1073/pnas.2135499100](https://doi.org/10.1073/pnas.2135499100)
- Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, Kim SK, Wall JD, Martin D, Jurka J, Milosavljevic A, de Jong PJ (2009) Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet* 5(6): e1000538
- Cardoso C, Leventer RJ, Ward HL, Toyo-oka K, Chung J, Gross A, Martin CL, Allanson J, Pilz DT, Olney AH (2003) Refinement of a 400-kb critical region allows genotypic differentiation between isolated lissencephaly, Miller-Dieker syndrome, and other phenotypes secondary to deletions of 17p13.3. *Am J Hum Genet* 72(4):918
- Carter NP (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 39:S16–S21
- Cassidy SB, Driscoll DJ (2009) Prader-Willi syndrome. *Eur J Hum Genet* 17(1):3–13
- Cassidy SB, Dykens E, Williams CA (2000) Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am J Med Genet* 97(2):136–146
- Chamberlain SJ, Lalande M (2010) Neurodevelopmental disorders involving genomic imprinting at human chromosome 15q11–q13. *Neurobiol Dis* 39(1):13–20
- Chen CP, Wang TH, Lin SP, Chern SR, Chen MR, Lee CC, Chen YJ, Wang W (2006) 24 Mb deletion of 6q22.1→q23.2 in an infant with pulmonary atresia, ventricular septal defect, microcephaly, developmental delay and facial dysmorphism. *Eur J Med Genet* 49 (6):516–519. doi:[10.1016/j.ejmg.2006.04.002](https://doi.org/10.1016/j.ejmg.2006.04.002)
- Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Genet* 17(2):154–163
- Christian SL, Fantes JA, Mewborn SK, Huang B, Ledbetter DH (1999) Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11–q13). *Hum Mol Genet* 8(6):1025–1037
- Cnossen MH, van der Est MN, Breuning MH, van Asperen CJ, Breslau-Siderius EJ, van der Ploeg AT, de Goede-Bolder A, van den Ouweland AMW, Halley DJJ, Niermeijer MF (1997) Deletions spanning the neurofibromatosis type 1 gene: Implications for genotype-phenotype correlations in neurofibromatosis type 1? *Hum Mutat* 9(5):458–464
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464(7289):704–712. doi:[10.1038/nature08516](https://doi.org/10.1038/nature08516)
- Consortium IMGSoA (1998) A full genome screen for autism with evidence for a linkage to a region on chromosome 7q. *Hum Mol Genet* 7:571–578
- Cook EH Jr, Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ, Lincoln A, Nix K, Haas R, Leventhal BL (1998) Linkage-disequilibrium mapping of autistic disorder, with 15q11–13 markers. *Am J Hum Genet* 62(5):1077

- Cook EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, Courchesne E (1997) Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am J Hum Genet* 60(4):928
- Curran ME, Atkinson DL, Ewart AK, Morris CA, Leppert MF, Keating MT (1993) The elastin gene is disrupted by a translocation associated with supravalvular aortic stenosis. *Cell* 73(1):159–168
- D'Angelo CS, Da Paz JA, Kim CA, Bertola DR, Castro CI, Varela MC, Koiffmann CP (2006) Prader-Willi-like phenotype: investigation of 1p36 deletion in 41 patients with delayed psychomotor development, hypotonia, obesity and/or hyperphagia, learning disabilities and behavioral problems. *Eur J Med Genet* 49(6):451–460. doi:[10.1016/j.ejmg.2006.02.001](https://doi.org/10.1016/j.ejmg.2006.02.001)
- D'Angelo CS, Gajecka M, Kim CA, Gentles AJ, Glotzbach CD, Shaffer LG, Koiffmann CP (2009) Further delineation of nonhomologous-based recombination and evidence for subtelomeric segmental duplications in 1p36 rearrangements. *Hum Genet* 125(5–6):551–563. doi:[10.1007/s00439-009-0650-9](https://doi.org/10.1007/s00439-009-0650-9)
- D'Angelo CS, Kohl I, Varela MC, de Castro CI, Kim CA, Bertola DR, Lourenco CM, Koiffmann CP (2010) Extending the phenotype of monosomy 1p36 syndrome and mapping of a critical region for obesity and hyperphagia. *Am J Med Genet A* 152A(1):102–110. doi:[10.1002/ajmg.a.33160](https://doi.org/10.1002/ajmg.a.33160)
- de Kovel CGF, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, von Spiczak S, Ostertag P (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 133(1):23–32
- De S, Teichmann SA, Babu MM (2009) The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome. *Genome Res* 19(5):785–794
- De Smith AJ, Purmann C, Walters RG, Ellis RJ, Holder SE, Van Haelst MM, Brady AF, Fairbrother UL, Dattani M, Keogh JM (2009) A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet* 18(17):3257–3265
- De Vries B, Pfundt R, Leisink M, Koolen DA, Vissers LELM, Janssen IM, Reijmersdal S, Nillesen WM, Huys EH, Leeuw N (2005) Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 77(4):606–616
- del Gaudio D, Fang P, Scaglia F, Ward PA, Craigen WJ, Glaze DG, Neul JL, Patel A, Lee JA, Irons M (2006) Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet Med* 8(12):784–792
- Ding F, Li HH, Zhang S, Solomon NM, Camper SA, Cohen P, Francke U (2008) SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3(3):e1709
- Ding F, Prints Y, Dhar MS, Johnson DK, Garnacho-Montero C, Nicholls RD, Francke U (2005) Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader-Willi syndrome mouse models. *Mamm Genome* 16(6):424–431
- Dobyns WB, Reiner O, Carrozzo R, Ledbetter DH (1993) Lissencephaly: a human brain malformation associated with deletion of the LIS1 gene located at chromosome 17p13. *JAMA* 270(23):2838–2842
- Doe CM, Relkovic D, Garfield AS, Dalley JW, Theobald DEH, Humby T, Wilkinson LS, Isles AR (2009) Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. *Hum Mol Genet* 18(12):2140–2148
- Doll A, Grzeschik KH (2001) Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams–Beuren syndrome. *Cytogenet Cell Genet* 95(1–2):20–27, 57012
- Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CAL, Dijkhuizen T, Bijlsma EK, Gijsbers ACJ, Hillhorst-Hofstee Y, Hordijk R, Verbruggen KT, Kerstjens-Frederikse W (2009) Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader–Willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet* 52(2–3):108–115
- Downs JA, Nussenzweig MC, Nussenzweig A (2007) Chromatin dynamics and the preservation of genetic information. *Nature* 447(7147):951–958. doi:[10.1038/nature05980](https://doi.org/10.1038/nature05980)

- Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, Thompson R, Traylor R, Bejjani BA, Shaffer LG, Rosenfeld JA, Lamb AN, Sahoo T (2011) Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. *Eur J Hum Genet* 18(11):1196–1201
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Kheradpour P, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Good PJ, Feingold EA, Crawford GE, Dekker J, Elinitzki L, Farnham PJ, Giddings MC, Gingeras TR, Guigo R, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Stamatoyannopoulos JA, Tennebaum SA, Weng Z, White KP, Wold B, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Eaton ML, Dobin A, Lassmann T, Tanzer A, Lagarde J, Lin W, Xue C, Williams BA, Zaleski C, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakraborty S, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Robyr D, Ruan X, Sammeth M, Sandu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Hayashizaki Y, Reymond A, Antonarakis SE, Hannon GJ, Ruan Y, Carninci P, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Grasfeder LL, Giresi PG, Battenhouse A, Sheffield NC, Showers KA, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniell RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Iyer VR, Sandhu KS, Zheng M, Wang P, Gertz J, Vielmetter J, Partridge EC, Varley KE, Gasper C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, Muratet MA, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Newberry JS, Levy SE, Absher DM, Wong WH, Blow MJ, Visel A, Pennachio LA, Elnitski L, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Davidson C, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Hunt T, Jungreis I, Kay M, Khurana E, Leng J, Lin MF, Loveland J, Lu Z, Manthavadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Sboner A, Searle S, Sisu C, Snow C, Steward C, Tapanan E, Tress ML, van Baren MJ, Washietl S, Wilming L, Zadissa A, Zhengdong Z, Brent M, Haussler D, Valencia A, Raymond A, Addleman N, Alexander RP, Auerbach RK, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harman A, Iyenger S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Larnar-Vincent N, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patascil D, Raha D, Ramirez L, Reed B, Shi M, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Struhl K, Weissman SM, Tenenbaum SA, Penalva LO, Karmakar S, Bhavania RR, Choudhury A, Domanus M, Ma L, Moran J, Victorson A, Auer T, Centarin L, Eichenlaub M, Gruhl F, Heerman S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Jain G, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen

- RS, Boatman L, Haugen E, Humbert R, Johnson AK, Johnson EM, Kutayavin TM, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sanchez ME, Sandstrom RS, Shafer AO, Stergachis AB, Thomas S, Vernot B, Vierstra J, Vong S, Weaver MA, Yan Y, Zhang M, Akey JA, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Stamatoyannopoulos JA, Beal K, Brazma A, Flicek P, Johnson N, Lusk M, Luscombe NM, Sobral D, Vaquerizas JM, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Miller W, Bickel PJ, Banfai B, Boley NP, Huang H, Li JJ, Noble WS, Bilmes JA, Buske OJ, Sahu AO, Kharchenko PV, Park PJ, Baker D, Taylor J, Lochovsky L (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74. doi:[10.1038/nature11247](https://doi.org/10.1038/nature11247)
- Earls LR, Fricke RG, Yu J, Berry RB, Baldwin LT, Zakharenko SS (2012) Age-dependent microRNA control of synaptic plasticity in 22q11 deletion syndrome and schizophrenia. *J Neurosci* 32(41):14132–14144. doi:[10.1523/JNEUROSCI.1312-12.2012](https://doi.org/10.1523/JNEUROSCI.1312-12.2012)
- El Tahir M, Kerr M, Jones R (2004) Two cases of generalized seizures and the Velocardiofacial syndrome—a clinically significant association? *J Intellect Disabil Res* 48(7):695–698
- Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296(5566):340–343. doi:[10.1126/science.1068996](https://doi.org/10.1126/science.1068996)
- Fang P, Lev-Lehman E, Tsai TF, Matsuura T, Benton CS, Sutcliffe JS, Christian SL, Kubota T, Halley DJ, Meijers-Heijboer H (1999) The Spectrum of Mutations in UBE3A Causing Angelman Syndrome. *Hum Mol Genet* 8(1):129–135
- Feng J, Schroer R, Yan J, Song W, Yang C, Bockholt A, Cook EH, Skinner C, Schwartz CE, Sommer SS (2006) High frequency of neurexin 1 β signal peptide structural variants in patients with autism. *Neurosci Lett* 409(1):10–13
- Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7(2):85–97
- Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. *Nat Genet* 18(2):168–170
- Forbes SH, Dorschner MO, Le R, Stephens K (2004) Genomic context of paralogous recombination hotspots mediating recurrent NF1 region microdeletion. *Genes Chromosomes Cancer* 41(1):12–25
- Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME (2006) Copy number variation: new insights in genome diversity. *Genome Res* 16(8):949–961
- Fulmer-Smentek SB, Francke U (2001) Association of acetylated histones with paternally expressed genes in the Prader–Willi deletion region. *Hum Mol Genet* 10(6):645–652
- Gajicka M, Saitta SC, Gentles AJ, Campbell L, Cipraro K, Geiger E, Catherwood A, Rosenfeld JA, Shaikh T, Shaffer LG (2010) Recurrent interstitial 1p36 deletions: evidence for germline mosaicism and complex rearrangement breakpoints. *Am J Med Genet A* 152A(12):3074–3083. doi:[10.1002/ajmg.a.33733](https://doi.org/10.1002/ajmg.a.33733)
- Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, Shafer N, Bernier R, Ferrero GB, Silengo M, Warren ST, Moreno CS, Fichera M, Romano C, Raskind WH, Eichler EE (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet* 7(11):e1002334. doi:[10.1371/journal.pgen.1002334](https://doi.org/10.1371/journal.pgen.1002334)
- Girirajan S, Vlangos CN, Szomju BB, Edelman E, Trevors CD, Dupuis L, Nizarati M, Bunyan DJ, Elsea SH (2006) Genotype–phenotype correlation in Smith–Magenis syndrome: evidence that multiple genes in 17p11.2 contribute to the clinical spectrum. *Genet Med* 8(7):417–427
- Glenn CC, Nicholls RD, Robinson WP, Saitoh S, Niikawa N, Schinzel A, Horsthemke B, Driscoll DJ (1993) Modification of 15q11–q13 DNA methylation imprints in unique Angelman and Prader–Willi patients. *Hum Mol Genet* 2(9):1377–1382

- Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, Reymond A, Sun M, Sawa A, Gusella JF, Kamiya A, Beckmann JS, Katsanis N (2012) KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 485 (7398):363–367. doi:10.1038/nature11091
- Grayton HM, Rujescu D, Collier DA (2012) Copy number variations in neurodevelopmental disorders. *Prog Neurobiol* 99:81–91
- Greenberg F, Guzzetta V, de Oca-Luna RM, Magenis R, Smith A, Richter S, Kondo I, Dobyns W, Patel P, Lupski J (1991) Molecular analysis of the Smith-Magenis syndrome: a possible contiguous-gene syndrome associated with del (17) (p11. 2). *Am J Hum Genet* 49(6):1207
- Greenberg F, Lewis RA, Potocki L, Glaze D, Parke J, Killian J, Murphy MA, Williamson D, Brown F, Dutton R (1996) Multi-disciplinary clinical study of Smith-Magenis syndrome (deletion 17p11. 2). *Am J Med Genet* 62(3):247–254
- Gregory RI, Randall TE, Johnson CA, Khosla S, Hatada I, O'Neill LP, Turner BM, Feil R (2001) DNA methylation is linked to deacetylation of histone H3, but not H4, on the imprinted genes *Snrpn* and *U2af1-rs1*. *Mol Cell Biol* 21(16):5426–5436
- Griffiths AJF, Gelbart WM, Miller JH (1999) Modern genetic analysis. W. H. Freeman, New York, Chromosomal rearrangements. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21367/>
- Gurling H, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A (2001) Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32. 2, 5q33. 2, and 8p21–22 and provides support for linkage to schizophrenia, on chromosomes 11q23. 3–24 and 20q12. 1–11.23. *Am J Hum Genet* 68(3):661–673
- Hatakeyama C, Anderson CL, Beever CL, Penaherrera MS, Brown CJ, Robinson WP (2004) The dynamics of X-inactivation skewing as women age. *Clin Genet* 66(4):327–332
- Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, De Kovel C, Baker C, Von Spiczak S (2009) 15q13. 3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* 41(2):160–162
- Hogart A, Leung KN, Wang NJ, Wu DJ, Driscoll J, Vallero RO, Schanen NC, LaSalle JM (2009) Chromosome 15q11–13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *J Med Genet* 46(2):86–93
- Hogart A, Wu D, LaSalle JM, Schanen NC (2008) The comorbidity of autism with the genomic disorders of chromosome 15q11.2–q13. *Neurobiol Dis* 38(2):181–191
- Hogart A, Wu D, LaSalle JM, Schanen NC (2010) The comorbidity of autism with the genomic disorders of chromosome 15q11. 2–q13. *Neurobiol Dis* 38(2):181–191
- Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* 37(1):31–40
- Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, Sciaky N, Dutton JW Jr, Lee HM, Chen X, Jin J, Bridges AS, Zylka MJ, Roth BL, Philpot BD (2012) Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* 481(7380):185–189. doi:10.1038/nature10726
- Huibregtse JM, Scheffner M, Beaudenon S, Howley PM (1995) A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci* 92 (7):2563–2567
- Hurles ME, Dermitzakis ET, Tyler-Smith C (2008) The functional impact of structural variation in humans. *Trends Genet* 24(5):238–245
- Hwu H, Liu C, Fann CSJ, Ou-Yang W, Lee SFC (2003) Linkage of schizophrenia with chromosome 1q loci in Taiwanese families. *Mol Psychiatry* 8(4):445–452
- IMGSAC (IMGSAC) (2001) A genome wide screen for autism: Strong evidence for linkage to chromosomes 2q, 7q and 16p. *Am J Hum Genet* 69:570–581
- Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE (2007) Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution. *Nat Genet* 39(11):1361–1368

- Johnson LW, Fishman RA, Schneider B, Parker F, Husson G, Webb W (1976) Familial supraaortic stenosis. Report of a large family and review of the literature. *Chest* 70 (4):494–500
- Johnstone KA, Dubose AJ, Futtner CR, Elmore MD, Brannan CI, Resnick JL (2005) A human imprinting centre demonstrates conserved acquisition but diverged maintenance of imprinting in a mouse model for Angelman syndrome imprinting defects. *Hum Mol Genet* 15(3):393–404
- Kernohan KD, Jiang Y, Tremblay DC, Bonvissuto AC, Eubanks JH, Mann MR, Berube NG (2010) ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. *Dev Cell* 18(2):191–202
- Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, Shen Y, Lally E, Weiss LA, Najm J, Kutsche K (2008) Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet* 82(1):199–207
- Kirk EP, Malaty-Brevaud V, Martini N, Lacoste C, Levy N, Maclean K, Davies L, Philip N, Badens C (2009) The clinical variability of the MECP2 duplication syndrome: description of two families with duplications excluding L1CAM and FLNA. *Clin Genet* 75(3):301–303
- Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, O'Donovan MC, Erdogan F, Owen MJ, Ropers HH (2008) Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum Mol Genet* 17(3):458–465
- Kishino T, Lalonde M, Wagstaff J (1997) UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* 15(1):70–73
- Kishore S, Stamm S (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311(5758):230–232
- Knoll JHM, Nicholls RD, Lalonde M (1989) On the parental origin of the deletion in Angelman syndrome. *Hum Genet* 83(2):205–206
- Knudsen GP, Neilson TC, Pedersen J, Kerr A, Schwartz M, Hulten M, Bailey ME, Orstavik KH (2006) Increased skewing of X chromosome inactivation in Rett syndrome patients and their mothers. *Eur J Hum Genet* 14(11):1189–1194
- Korf BR (2002) Clinical Features and pathobiology of neurofibromatosis 1. *J Child Neurol* 17 (8):573–577
- Kosak ST, Groudine M (2004) Form follows function: the genomic organization of cellular differentiation. *Genes Dev* 18(12):1371–1384. doi:[10.1101/gad.1209304](https://doi.org/10.1101/gad.1209304)
- Lai CSL, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED (2000) The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *Am J Hum Genet* 67(2):357–368
- Lalonde M, Calciano M (2007) Molecular epigenetics of Angelman syndrome. *Cell Mol Life Sci* 64(7):947–960
- LaSalle JM (2011) A genomic point-of-view on environmental factors influencing the human brain methylome. *Epigenetics* 6(7):862–869
- Lee JA, Lupski JR (2006) Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* 52(1):103
- Leung KN, Vallero RO, DuBose AJ, Resnick JL, LaSalle JM (2009) Imprinting regulates mammalian snoRNA-encoding chromatin decondensation and neuronal nucleolar size. *Hum Mol Genet* 18(22):4227–4238
- Li J, Harris RA, Cheung SW, Coarfa C, Jeong M, Goodell MA, White LD, Patel A, Kang SH, Shaw C, Chinault AC, Gambin T, Gambin A, Lupski JR, Milosavljevic A (2012) Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. *PLoS Genet* 8(5):e1002692. doi:[10.1371/journal.pgen.1002692](https://doi.org/10.1371/journal.pgen.1002692)
- Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322

- López-Correa C, Dorschner M, Brems H, Lázaro C, Clementi M, Upadhyaya M, Dooijes D, Moog U, Kehrer-Sawatzki H, Rutkowski JL (2001) Recombination hotspot in NF1 microdeletion patients. *Hum Mol Genet* 10(13):1387–1392
- Lossie A, Whitney M, Amidon D, Dong H, Chen P, Theriaque D, Hutson A, Nicholls R, Zori R, Williams C (2001) Distinct phenotypes distinguish the molecular classes of Angelman syndrome. *J Med Genet* 38(12):834–845
- Lugtenberg D, Kleefstra T, Oudakker AR, Nillesen WM, Yntema HG, Tzschach A, Raynaud M, Rating D, Journal H, Chelly J, Goizet C, Lacombe D, Pedespan JM, Echenne B, Tariverdian G, O'Rourke D, King MD, Green A, van Kogelenberg M, Van Esch H, Geçz J, Hamel BC, van Bokhoven H, de Brouwer AP (2009) Structural variation in Xq28: MECP2 duplications in 1 % of patients with unexplained XLMR and in 2 % of male patients with severe encephalopathy. *Eur J Hum Genet* 17(4):444–453
- Lupski JR (2006) Genome structural variation and sporadic disease traits. *Nat Genet* 38(9):974
- Lupski JR, Stankiewicz P (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* 1(6):e49
- Lynch DR, McDonald-McGinn DM, Zackai EH, Emanuel BS, Driscoll DA, Whitaker LA, Fischbeck KH (1995) Cerebellar atrophy in a patient with velocardiofacial syndrome. *J Med Genet* 32(7):561–563
- Makedonski K, Abuhazira L, Kaufman Y, Razin A, Shemer R (2005) MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects UBE3A expression. *Hum Mol Genet* 14(8):1049–1058
- Makoff AJ, Flomen RH (2007) Detailed analysis of 15q11–q14 sequence corrects errors and gaps in the public access sequence to fully reveal large segmental duplications at breakpoints for Prader-Willi, Angelman, and inv dup (15) syndromes. *Genome Biol* 8(6):R114
- Malcolm S, Clayton-Smith J, Nichols M, Pembrey M, Armour J, Jeffreys A, Robb S, Webb T (1991) Uniparental paternal disomy in Angelman's syndrome. *Lancet* 337(8743):694–697
- Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? *Curr Opin Genet Dev* 15(2):163–176
- Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang Y, Benton CS, Rommens JM, Beaudet AL (1997) De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat Genet* 15(1):74–77
- McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PIW, Maller JB, Kirby A (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 40(10):1166–1174
- McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O (2009) Microduplications of 16p11. 2 are associated with schizophrenia. *Nat Genet* 41(11):1223–1227
- Meechan DW, Tucker ES, Maynard TM, LaMantia AS (2009) Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 deletion/DiGeorge syndrome. *Proc Natl Acad Sci* 106(38):16434–16445
- Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K, Huang S, Maloney VK, Crolla JA, Baralle D (2008) Recurrent rearrangements of chromosome 1q21. 1 and variable pediatric phenotypes. *N Engl J Med* 359(16):1685–1699
- Meguro-Horike M, Yasui DH, Powell W, Schroeder DI, Oshimura M, LaSalle JM, Horike S (2011a) Neuron-specific impairment of inter-chromosomal pairing and transcription in a novel model of human 15q-duplication syndrome. *Hum Mol Genet* 20(19):3798–3810. doi:[10.1093/hmg/ddr298](https://doi.org/10.1093/hmg/ddr298)
- Meguro-Horike M, Yasui DH, Powell W, Schroeder DI, Oshimura M, LaSalle JM, Horike S (2011b) Neuron-specific impairment of inter-chromosomal pairing and transcription in a novel model of human 15q-duplication syndrome. *Hum Mol Genet* 20(19):3798–3810
- Meins M, Lehmann J, Gerresheim F, Herchenbach J, Hagedorn M, Hameister K, Epplen J (2005) Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. *J Med Genet* 42(2):e12

- Merla G, Ucla C, Guipponi M, Reymond A (2002) Identification of additional transcripts in the Williams–Beuren syndrome critical region. *Hum Genet* 110(5):429–438. doi:[10.1007/s00439-002-0710-x](https://doi.org/10.1007/s00439-002-0710-x)
- Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C, Cox GF, Dickinson H, Gentile J, Harris DJ (2009) Microdeletion/duplication at 15q13. 2q13. 3 among individuals with features of autism and other neuropsychiatric disorders. *J Med Genet* 46(4):242–248
- Mitchell MM, Woods R, Chi LH, Schmidt RJ, Pessah IN, Kostyniak PJ, Lasalle JM (2012a) Levels of select PCB and PBDE congeners in human postmortem brain reveal possible environmental involvement in 15q11–q13 duplication autism spectrum disorder. *Environ Mol Mutagen* 53:589–598. doi:[10.1002/em.21722](https://doi.org/10.1002/em.21722)
- Mitchell MM, Woods R, Chi LH, Schmidt RJ, Pessah IN, Kostyniak PJ, Lasalle JM (2012b) Levels of select PCB and PBDE congeners in human postmortem brain reveal possible environmental involvement in 15q11–q13 duplication autism spectrum disorder. *Environ Mol Mutagen* 53(8):589–598. doi:[10.1002/em.21722](https://doi.org/10.1002/em.21722)
- Moeschler JB, Mohandas T, Hawk AB, Noll WW (2002) Estimate of prevalence of proximal 15q duplication syndrome. *Am J Med Genet* 111(4):440–442
- Moog U, Smeets EEJ, van Roozendaal KEP, Schoenmakers S, Herbergs J, Schoonbrood-Lenssen AMJ, Schrandt-Stumpel CTRM (2003) Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP2). *Eur J Paediatr Neurol* 7(1):5–12
- Moreno-De-Luca D, Sanders S, Willsey A, Mulle J, Lowe J, Geschwind D (2012) Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. *Mol Psychiatry*, Oct 9. doi:[10.1038/mp.2012.138](https://doi.org/10.1038/mp.2012.138). [Epub ahead of print]
- Morison IM, Ramsay JP, Spencer HG (2005) A census of mammalian imprinting. *Trends Genet* 21(8):457–465
- Moss EM, Batshaw ML, Solot CB, Gerdes M, McDonald-McGinn DM, Driscoll DA, Emanuel BS, Zackai EH, Wang PP (1999) Psychoeducational profile of the 22q11. 2 microdeletion: a complex pattern. *J Pediatr* 134(2):193–198
- Mulle JG, Dodd AF, McGrath JA, Wolyniec PS, Mitchell AA, Shetty AC, Sobreira NL, Valle D, Rudd MK, Satten G (2010) Microdeletions of 3q29 confer high risk for schizophrenia. *Am J Hum Genet* 87(2):229–236
- Murphy K, Owen M (2001) Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. *Br J Psychiatry* 179(5):397–402
- Murthy S, Nygren A, El Shakankiry H, Schouten J, Al Khayat A, Ridha A, Al Ali M (2007) Detection of a novel familial deletion of four genes between BP1 and BP2 of the Prader-Willi/Angelman syndrome critical region by oligo-array CGH in a child with neurological disorder and speech impairment. *Cytogenet Genome Res* 116(1–2):135–140
- Nguyen DQ, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. *PLoS Genet* 2(2):e20
- Ni X, Valente J, Azevedo MH, Pato MT, Pato CN, Kennedy JL (2007) Connexin 50 gene on human chromosome 1q21 is associated with schizophrenia in matched case–control and family-based studies. *J Med Genet* 44(8):532–536
- Nischal S, Bhattacharyya S, Christopheit M, Yu Y, Zhou L, Bhagat T, Sohal D, Will B, Mo Y, Suzuki M, Pardanani A, McDevitt M, Maciejewski JP, Melnick AM, Grealley J, Steidl U, Moliterno A, Verma A (2012) Methylome profiling reveals distinct alterations in phenotypic and mutational subgroups of myeloproliferative neoplasms. *Cancer Res* 73:1076–1085. doi:[10.1158/0008-5472.CAN-12-0735](https://doi.org/10.1158/0008-5472.CAN-12-0735)
- Payer B, Lee JT (2008) X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* 42:733–772
- Paylor R, Glaser B, Mupo A, Ataliotis P, Spencer C, Sobotka A, Sparks C, Choi CH, Oghalai J, Curran S (2006) Tbx1 haploinsufficiency is linked to behavioral disorders in mice and humans: implications for 22q11 deletion syndrome. *Proc Natl Acad Sci* 103(20):7729–7734
- Penner KA, Johnston J, Faircloth BH, Irish P, Williams CA (1993) Communication, cognition, and social interaction in the Angelman syndrome. *Am J Med Genet* 46(1):34–39

- Peoples RJ, Cisco MJ, Kaplan P, Francke U (1998) Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams–Beuren syndrome deletion at 7q11.23. *Cytogenet Cell Genet* 82(3–4):238–246
- Perk J, Makedonski K, Lande L, Cedar H, Razin A, Shemer R (2002) The imprinting mechanism of the Prader-Willi/Angelman regional control center. *EMBO J* 21(21):5807–5814
- Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurler ME, Tyler-Smith C (2008) Copy number variation and evolution in humans and chimpanzees. *Genome Res* 18(11):1698–1710
- Pinto D, Marshall C, Feuk L, Scherer SW (2007) Copy-number variation in control population cohorts. *Hum Mol Genet* 16(R2):R168–R173
- Pober BR (2010) Williams–Beuren syndrome. *N Engl J Med* 362(3):239–252
- Prescott TE, Rodningen OK, Bjornstad A, Stray-Pedersen A (2009) Two brothers with a microduplication including the MECP2 gene: rapid head growth in infancy and resolution of susceptibility to infection. *Clin Dysmorphol* 18(2):78–82
- Rasmussen SA, Colman SD, Ho VT, Abernathy CR, Arn PH, Weiss L, Schwartz C, Saul RA, Wallace MR (1998) Constitutional and mosaic large NF1 gene deletions in neurofibromatosis type 1. *J Med Genet* 35(6):468–471
- Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP (2009) A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* 106(3):671–678
- Ravn K, Nielsen JB, Skjeldal OH, Kerr A, Hulten M, Schwartz M (2005) Large genomic rearrangements in MECP2. *Hum Mutat* 25(3):324
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444–454
- Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2(1):21–32
- Reis A, Dittrich B, Greger V, Buiting K, Lalonde M, Gillessen-Kaesbach G, Anvret M, Horsthemke B (1994) Imprinting mutations suggested by abnormal DNA methylation patterns in familial Angelman and Prader-Willi syndromes. *Am J Hum Genet* 54(5):741–747
- Repetto GM, White LM, Bader PJ, Johnson D, Knoll JHM (1998) Interstitial duplications of chromosome region. *Am J Med Genet* 79:82–89
- Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama–Kurtycz CM, Hartley SL, Sprague RE, Heintzelman AT, Shriberg LD (2012) Phenotype of FOXP2 haploinsufficiency in a mother and son. *Am J Med Genet A* 158A(1):174–181
- Robinson W, Dutly F, Nicholls R, Bernasconi F, Penaherrera M, Michaelis R, Abeliovich D, Schinzel A (1998) The mechanisms involved in formation of deletions and duplications of 15q11–q13. *J Med Genet* 35(2):130–136
- Robinson WP, Beever C, Brown CJ, Stephenson MD (2001) Skewed X inactivation and recurrent spontaneous abortion. *Semin Reprod Med* 19(2):175–181
- Rosenfeld JA, Amrom D, Andermann E, Andermann F, Veilleux M, Curry C, Fisher J, Deputy S, Aylsworth AS, Powell CM, Manickam K, Heese B, Maisenbacher M, Stevens C, Ellison JW, Upton S, Moeschler J, Torres-Martinez W, Stevens A, Marion R, Pereira EM, Babcock M, Morrow B, Sahoo T, Lamb AN, Ballif BC, Paciorkowski AR, Shaffer LG (2012) Genotype-phenotype correlation in interstitial 6q deletions: a report of 12 new cases. *Neurogenetics* 13(1):31–47. doi:[10.1007/s10048-011-0306-5](https://doi.org/10.1007/s10048-011-0306-5)
- Rosenfeld JA, Crolla JA, Tomkins S, Bader P, Morrow B, Gorski J, Troxell R, Forster-Gibson C, Cilliers D, Hislop RG, Lamb A, Torchia B, Ballif BC, Shaffer LG (2010) Refinement of causative genes in monosomy 1p36 through clinical and molecular cytogenetic characterization of small interstitial deletions. *Am J Med Genet A* 152A(8):1951–1959. doi:[10.1002/ajmg.a.33516](https://doi.org/10.1002/ajmg.a.33516)
- Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalonde M (1998) An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat Genet* 19(1):15–16

- Rujescu D, Ingason A, Cichon S, Pietiläinen OPH, Barnes MR, Touloupoulou T, Picchioni M, Vassos E, Ettinger U, Bramon E (2009) Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 18(5):988–996
- Sahoo T, Del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, Garnica A, Cheung SW, Beaudet AL (2008) Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 40(6):719–721
- Schaschl H, Aitman T, Vyse T (2009) Copy number variation in the human genome and its implication in autoimmunity. *Clin Exp Immunol* 156(1):12–16
- Schroeder DI, Lott P, Korf I, LaSalle JM (2011) Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res* 21(10):1583–1591. doi:[10.1101/gr.119131.110](https://doi.org/10.1101/gr.119131.110)
- Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D (1998) Autism and maternally derived aberrations of chromosome 15q. *Am J Med Genet* 76(4):327–336
- Scoles HA, Urraca N, Chadwick SW, Reiter LT, LaSalle JM (2011) Increased copy number for methylated maternal 15q duplications leads to changes in gene and protein expression in human cortical samples. *Mol Autism* 2(1):19. doi:[10.1186/2040-2392-2-19](https://doi.org/10.1186/2040-2392-2-19)
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M (2007) Strong association of de novo copy number mutations with autism. *Science* 316(5823):445–449
- Shapira SK, McCaskill C, Northrup H, Spikes AS, Elder FF, Sutton VR, Korenberg JR, Greenberg F, Shaffer LG (1997) Chromosome 1p36 deletions: the clinical phenotype and molecular characterization of a common newly delineated syndrome. *Am J Hum Genet* 61(3):642–650. doi:[10.1086/515520](https://doi.org/10.1086/515520)
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38(9):1038–1042
- Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE, Schroer RJ, Novara F, De Gregori M, Ciccone R (2008) A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat Genet* 40(3):322–328
- Shaw CJ, Withers MA, Lupski JR (2004) Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. *Am J Hum Genet* 75(1):75–81
- Shinawi M, Liu P, Kang SHL, Shen J, Belmont JW, Scott DA, Probst FJ, Craigen WJ, Graham BH, Pursley A (2010) Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 47(5):332–341
- Skryabin BV, Gubar LV, Seeger B, Pfeiffer J, Handel S, Robeck T, Karpova E, Rozhdestvensky TS, Brosius J (2007) Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet* 3(12):e235
- Soejima H, Wagstaff J (2005) Imprinting centers, chromatin structure, and disease. *J Cell Biochem* 95(2):226–233
- Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455(7210):232–236
- Tang MH, Varadan V, Kamalakaran S, Zhang MQ, Dimitrova N, Hicks J (2012) Major chromosomal breakpoint intervals in breast cancer co-localize with differentially methylated regions. *Front Oncol* 2:197. doi:[10.3389/fonc.2012.00197](https://doi.org/10.3389/fonc.2012.00197)

- Thatcher K, Peddada S, Yasui D, LaSalle JM (2005) Homologous pairing of 15q11–13 imprinted domains in brain is developmentally regulated but deficient in Rett and autism samples. *Hum Mol Genet* 14:785–797
- TISC (ISC) (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455(7210):237–241, http://www.nature.com/nature/journal/v455/n7210/supinfo/nature07239_S1.html
- Tolosa A, Sanjuan J, Dagnall AM, Molto MD, Herrero N, de Frutos R (2010) FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. *BMC Med Genet* 11:114. doi:10.1186/1471-2350-11-114
- Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K, Lugtenberg D, Bienvenu T, Jensen LR, Géczy J (2005) Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am J Hum Genet* 77(3):442
- Velinov M, Novelli A, Gu H, Fenko M, Dolzhanskaya N, Bernardini L, Capalbo A, Dallapiccola B, Jenkins EC, Brown WT (2009) De-novo 2.15 Mb terminal Xq duplication involving MECP2 but not L1CAM gene in a male patient with mental retardation. *Clin Dysmorphol* 18(1):9–12
- Vitali P, Royo H, Marty V, Bortolin-Cavaillé ML, Cavaillé J (2010) Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J Cell Sci* 123(1):70–83
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320(5875):539
- Wandstrat A, Leana-Cox J, Jenkins L, Schwartz S (1998) Molecular cytogenetic evidence for a common breakpoint in the largest inverted duplications of chromosome 15. *Am J Hum Genet* 62(4):925–936
- Wang NJ, Liu D, Parokony AS, Schanen NC (2004a) High-resolution molecular characterization of 15q11–q13 rearrangements by array comparative genomic hybridization (array CGH) with detection of gene dosage. *Am J Hum Genet* 75(2):267
- Wang Y, Fischle W, Cheung W, Jacobs S, Khorasanizadeh S, Allis CD (2004b) Beyond the double helix: writing and reading the histone code. *Novartis Found Symp* 259:3–17, discussion 17–21, 163–169
- Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Saemundsen E, Stefansson H, Ferreira MAR, Green T (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358(7):667–675
- Willatt L, Cox J, Barber J, Cabanas ED, Collins A, Donnai D, FitzPatrick DR, Maher E, Martin H, Parnau J (2005) 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am J Hum Genet* 77(1):154–160
- Williams NM, Glaser B, Norton N, Williams H, Pierce T, Moskvina V, Monks S, Del Favero J, Goossens D, Rujescu D (2008) Strong evidence that GNB1L is associated with schizophrenia. *Hum Mol Genet* 17(4):555–566
- Xie H, Wang M, Bonaldo Mde F, Smith C, Rajaram V, Goldman S, Tomita T, Soares MB (2009) High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum. *Nucleic Acids Res* 37(13):4331–4340
- Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T, Niikawa N, Ogawa M, Wagstaff J, Kishino T (2003) Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. *Hum Mol Genet* 12(8):837–847
- Yasui DH, Scoles HA, Horike S, Meguro-Horike M, Dunaway KW, Schroeder DI, LaSalle JM (2011) 15q11.2–13.3 chromatin analysis reveals epigenetic regulation of CHRNA7 with deficiencies in Rett and autism brain. *Hum Mol Genet* 20(22):4311–4323. doi:10.1093/hmg/ddr357
- Young JM, Endicott RLM, Parghi SS, Walker M, Kidd JM, Trask BJ (2008) Extensive copy-number variation of the human olfactory receptor gene family. *Am J Hum Genet* 83(2):228

- Zahir FR, Baross A, Delaney AD, Eyedoux P, Fernandes ND, Pugh T, Marra MA, Friedman JM (2008) A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1 α . *J Med Genet* 45(4):239–243
- Zepeda-Mendoza CJ, Lemus T, Yanez O, Garcia D, Valle-Garcia D, Meza-Sosa KF, Gutierrez-Arcelus M, Marquez-Ortiz Y, Dominguez-Vidana R, Gonzaga-Jauregui C, Flores M, Palacios R (2010) Identical repeated backbone of the human genome. *BMC Genomics* 11:60. doi:[10.1186/1471-2164-11-60](https://doi.org/10.1186/1471-2164-11-60)
- Zhang Z, Tang H, Wang Z, Zhang B, Liu W, Lu H, Xiao L, Liu X, Wang R, Li X, Wu M, Li G (2011) MiR-185 targets the DNA methyltransferases 1 and regulates global DNA methylation in human glioma. *Mol Cancer* 10:124. doi:[10.1186/1476-4598-10-124](https://doi.org/10.1186/1476-4598-10-124)
- Zheng Y, Wang X, Gu N, Feng G, Zou F, Qin W, Zhang J, Lin W, Tao R, Qian X (2006) A two-stage linkage analysis of Chinese schizophrenia pedigrees in 10 target chromosomes. *Biochem Biophys Res Commun* 342(4):1049–1057

Chapter 8

Impact of the Early-Life Environment on the Epigenome and Behavioral Development

Benoit Labonté and Gustavo Turecki

Abstract The environment in which we live, and especially the early-life environment, regulates our behavioral development. Adversity during early life is strongly associated with problems in behavioral regulation and psychopathology in adulthood. Until recently, the mechanisms responsible for behavioral changes induced by early-life adversity were not clear. However, recent evidence suggests that early-life environment induces behavioral changes through epigenetic mechanisms controlling the expression of genes involved in the regulation of behavior. Thus, the epigenome mediates the effects of environmental variability on behavioral, physiological, and pathological responses increasing vulnerability toward suicidal behaviors. Numerous findings in animals and humans support this view. This chapter reviews the evidence suggesting that epigenetic changes are induced by the early environment and impact the regulation of gene expression in the brain increasing the risk for suicidal behaviors.

8.1 The Burden of Early-Life Adversity and Suicide

Children in our society are all too often subjected to maltreatment, which is frequently perpetrated by their parents or caregivers. Accordingly, childhood maltreatment is a global problem of significant proportion that affects children of all ages, race, economic, and cultural backgrounds (Children'sBureau 2010). There are four main types of childhood maltreatment; these are sexual abuse, physical abuse, psychological abuse, as well as parental neglect (Gilbert et al. 2009). With more than three million reports of child maltreatment in the USA in 2009 and similar statistics elsewhere in Western societies (Children'sBureau 2010), early-life

B. Labonté • G. Turecki, M.D., Ph.D. (✉)

McGill Group for Suicide Studies, Department of Psychiatry, Douglas Mental Health Institute, 6875 LaSalle Blvd, Montreal, QC, Canada H4H 1R3

e-mail: benoit.labonte@mail.mcgill.ca; gustavo.turecki@mcgill.ca

adversity represents one of the major risk factors associated with higher prevalence of suicidal behaviors (Evans et al. 2005; Santa Mina and Gallop 1998). From an epidemiological point of view, trauma exposure in children is estimated to range between 25 and 45 %, although the rates reported vary considerably between studies and according to the definition of abuse types (Scher et al. 2004; Briere and Elliott 2003; Gorey and Leslie 1997; McCauley et al. 1997; Heim et al. 2010).

The economic burden of child maltreatment and trauma resides mainly in its impact on the development of psychopathology later during adulthood. Indeed, child trauma, and particularly child sexual and physical abuse (CSA and CPA), is associated with increased risks of psychiatric disorders including depression, anxiety, bipolar disorder, substance abuse, and suicide (Evans et al. 2005; Molnar et al. 2001; Santa Mina and Gallop 1998; Heim and Nemeroff 2001; Kendler et al. 2000, 2004; Kaplan and Klinetob 2000; Agid et al. 1999; Fergusson et al. 1996). In addition to increasing the risk of psychiatric disorders, CSA and CPA also associate with earlier age of onset of psychopathology, chronic course, more severe outcomes, poorer recovery rates, and more importantly, with a 12 times higher odds of suicidal behaviors (Dinwiddie et al. 2000; Gladstone et al. 2004; Jaffee et al. 2002; Zlotnick et al. 2001; Brown and Moran 1994; Tanskanen et al. 2004; Bensley et al. 1999; Molnar et al. 2001).

8.2 Epigenetic Consequences of Early-Life Adversity on the Brain

Early-life adversity is frequently associated with maladaptive patterns of behavioral responses often leading to pervasive interpersonal difficulties, enhanced reactivity to stress, and increased risk of psychopathology. While substantial theoretical and empirical work supports the relationship between childhood adversity and development of negative mental health outcomes in adulthood, the critical question has been what molecular processes mediate these associations. In other words: “What long-lasting molecular mechanisms take place as a result of the adverse life experience that could be associated with increased risk for psychopathology?” Despite the complexity of this question, this chapter reviews the evidence suggesting that molecular alterations result from variation in early-life environment through epigenetic processes that modulate behaviors in animal models and increase the risk for suicide in humans.

Although the same DNA is found in every cell of our body, cells differentiate into specific cell types and synthesize different proteins, allowing them to evolve and adapt to specific environments. This whole process is believed to involve epigenetic mechanisms.

Epigenetics refers to the study of the epigenome: chemical modifications taking place in or around the DNA molecule and altering the capacity of a gene to be activated and to produce the mRNA it encodes. There are several epigenetic mechanisms, including histone modifications (Kouzarides 2007), DNA methylation

(Klose and Bird 2006), and hydroxymethylation (Kriaucionis and Heintz 2009). Another mechanism, which is not a priori an epigenetic mechanism but is often classified as such, is the posttranslational regulation of gene expression via microRNA (He and Hannon 2004). Given the high complexity of DNA organization, these modifications are expected to follow defined patterns allowing the underlying molecular mechanisms to be performed correctly and to decode DNA in the context of chromatin.

Within the cells' nuclei, DNA is packaged into a structure called chromatin, composed of nucleosomes, the fundamental unit of chromatin, around which 146 bp of DNA is wrapped. The nucleosome itself is formed by an octamer of the four core histones (H2A, H2B, H3, and H4), globular structures with a tail of amino acids which can be modified by the addition or the removal of chemical residues. Chromatin has two structures: euchromatin, the active state associated with gene transcription, and its counterpart, heterochromatin, the inactive state corresponding to gene repression.

The chromatin state is dynamically regulated by the recruitment of proteins carrying intrinsic enzymatic activity to histone modifications (Clements et al. 2003; Fischle et al. 2005; Nelson et al. 2006; Pray-Grant et al. 2005; Santos-Rosa et al. 2003) which induce the opening or the closing of chromatin (Wysocka et al. 2005, 2006). Up to eight types of histone modifications have been characterized (methylation (lysine, arginine), acetylation, phosphorylation, ubiquitylation, sumoylation, deimination, ADP ribosylation, and proline isomerization), although they might not be all found in eukaryotic cells (Kouzarides 2007). While these marks may contribute concomitantly to regulate gene expression, most attention has been focused on lysine methylation and acetylation. For instance, methylation at specific lysines (K) of the third histone such as the fourth (H; H3K4), 36th (H3K36), and H3K79 (Kirmizis et al. 2007; Salcedo-Amaya et al. 2009; Barrera et al. 2008; Pokholok et al. 2005; Wang et al. 2008; Xiao et al. 2007) has been associated with active transcription while methylation at H3K9, H3K27, and H4K20 often correlates with transcriptional repression (Wang et al. 2008; Barski et al. 2007; Bannister et al. 2001; Botuyan et al. 2006; Lan et al. 2007; Nielsen et al. 2001; Sanders et al. 2004; Swigut and Wysocka 2007). It is likely that each of these modifications has a distinct signature profile which may overlap to form a histone code controlling the structure of the chromatin, as well as gene transcription, according to cell needs. However, this code is still far from being cracked or understood.

DNA methylation is a posttranscriptional modification that refers mainly to the transfer of a methyl group (CH_3) from an *S*-adenosyl-L-methionine (AdoMet) donor to the 5' carbon of the cytosine from dinucleotide CpG sequences. This process requires the enzymatic activity of DNA methyltransferase (DNMT) proteins among which DNMT3a and DNMT3b are called *de novo* methylases because they introduce methyl groups at previously unmethylated cytosines (Hata et al. 2002; Okano et al. 1998). In contrast, DNA hydroxymethylation refers to the oxidation of preexisting 5' methylcytosine to 5' hydroxymethylation by enzymes by the TET family (Tahiliani et al. 2009; Ito et al. 2010).

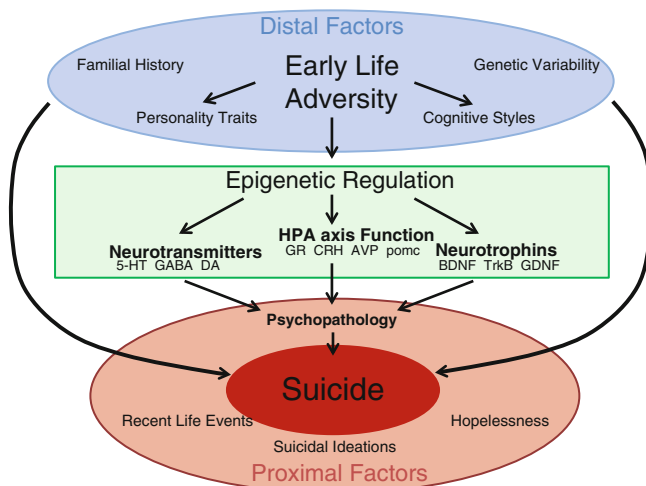


Fig. 8.1 The postulated relationships between early-life adversity, epigenetic regulation, and psychopathology

In somatic cells, approximately 80 % of CpGs are methylated (Tucker 2001). The remaining unmethylated ones are grouped in CpG-enriched regions called CpG islands, often found in the 5' regulatory regions of genes. Thus, DNA methylation in gene promoters has classically been associated with translational repression by interfering with the recruitment and the binding of the transcriptional machinery to gene regulatory regions (Klose and Bird 2006). However, when DNA methylation is found within the gene body, it has been associated with transcriptional activation and alternative transcript selection (Maunakea et al. 2010). In contrast, recent findings suggest that hydroxymethylation is enriched in the gene body of active genes (Mellen et al. 2012). Moreover, different cell types exhibit distinct methylation and hydroxymethylation patterns that confer a specificity of expression based on the requirements of each cell type (Mellen et al. 2012; Iwamoto et al. 2011).

In the following sections, we review findings in gene systems that have been targeted by studies investigating epigenetic factors associated with the social environment, specifically in the context of suicide. We review animal studies using models of early-life environmental variation, and studies focusing on suicide emphasizing those investigating the effect of early-life adversity. Overall we review the evidence suggesting that epigenetic mechanisms may be involved in the modification of gene expression induced by environmental factors. While most individuals who die by suicide do not have a history of abuse during childhood, a significant minority does, and in this subgroup, the association is very strong. Among the systems reviewed in this chapter, we focus in particular on gene systems coding for components of the hypothalamus-pituitary-adrenal (HPA) axis and related signaling hormones and molecules, as well as neurotrophic factors, their receptors, and neurotransmitters. Figure 8.1 displays the postulated relationships between early-life adversity, epigenetic regulation, and psychopathology.

8.3 HPA Axis Alterations Induced by Early-Life Adversity

Child abuse has been proposed to induce its long-term behavioral consequences partly by altering the neural circuits involved in the regulation of stress (Heim et al. 2008b). The HPA axis is the main stress regulatory system (Pariante and Lightman 2008). Under stressful conditions, corticotropin-releasing factor (CRF) and vasopressin (AVP) are released from the hypothalamus. CRF and AVP induce the release of adreno corticotropic hormone (ACTH) and pro-opiomelanocortin (POMC) from the pituitary gland to the blood which then travels to the adrenal cortex where they induce the release of glucocorticoids—cortisol in humans and corticosterone in rodents—to the blood. Glucocorticoids then act at each level of the HPA axis to decrease the release of CRF, AVP, POMC, and ACTH and regulate the stress response. While the HPA axis can be regulated at different levels, the main locus of regulation lies in the hippocampus where glucocorticoids bind glucocorticoid receptors (GR) and induce an inhibitory feedback on the activation of the HPA axis to bring the activity of the stress response back to basal levels.

From a structural point of view, childhood abuse and neglect have been associated with volume loss in the hippocampus (Bremner et al. 1997; Stein et al. 1997; Driessen et al. 2000), altered cortical symmetry in the frontal lobe (Carrion et al. 2001) and superior temporal gyrus (de Bellis et al. 2002), as well as a reduced neuronal density and/or neuronal integrity in the anterior cingulate gyrus (de Bellis et al. 2002). One study also reported poorer hippocampal activation on a memory task in patients with a history of childhood abuse (Bremner et al. 2003). Importantly, the structural consequences of child abuse are thought to be time dependent, implying that particular brain regions may have unique windows of vulnerability to the effects of child abuse (Andersen et al. 2008).

From a molecular point of view, depressed patients with a history of child abuse have been reported to exhibit higher ACTH and cortisol levels following stress and dexamethasone (DEX) challenges (Heim et al. 2000, 2008a). Interestingly, in these studies, both ACTH and cortisol levels did not differ significantly between depressed subjects without a history of childhood abuse and controls (Heim et al. 2000, 2008a). Childhood abuse, and particularly physical abuse, has also been shown to increase corticotropin-releasing hormone (CRH) levels (Heim et al. 2008b; Carpenter et al. 2004) and to decrease cerebral spinal fluid oxytocin levels (Heim et al. 2009). More recently, low hippocampal GR levels have been reported in suicide completers with a history of childhood abuse (McGowan et al. 2009; Labonte et al. 2012b) but not in non-abused suicide completers. Altogether, these alterations are believed to lead to important behavioral changes that may increase the predisposition toward suicidal behavior later in life.

This work is also substantiated with findings from animal work. For instance, the development of the HPA axis has been shown to be modulated by maternal behavior in rats. Depressive-like behaviors (Francis et al. 1999) associated with altered HPA axis feedback (Liu et al. 1997) and low GR mRNA hippocampal levels (Liu et al. 1997) are common features in rats raised by mothers providing maternal

care defined by low levels of licking and grooming (LG). A comprehensive model involving modulation at numerous levels, including hormonal, synaptic, and molecular changes, has been proposed in an attempt to characterize the molecular pathways involved in the modulatory effects of maternal behavior in rats. High maternal LG levels, which can also be mimicked by handling pups during early life, induce a physiological response involving the release of thyroid hormone plasma levels. This increases 5-HT activity in the raphe nuclei, and consequently stimulates serotonin (5-HT) turnover in the hippocampus and frontal cortex (Mitchell et al. 1990; Smythe et al. 1994; Meaney et al. 1987). Via activation of the G-protein-coupled 5-HT₇ receptor (Laplanche et al. 2002), it is believed that 5-HT activates a c-AMP/PKA-dependant intracellular cascade increasing the expression of nerve growth factor 1-A (NGFI-A) and activator protein-2 (AP-2) in the hippocampus (Meaney et al. 2000). NGFI-A and AP-2 are activating transcription factors with putative binding sites within the GR promoter region (McCormick et al. 2000) that increase GR mRNA levels in the hippocampus of the offspring. This complex process is attenuated in rats raised by low LG mothers according to the molecular and behavioral processes mentioned previously and resulting in relatively lower GR expression in the hippocampus (Weaver et al. 2004; McGowan et al. 2011). Interestingly, most of these regulatory changes are temporally stable and are maintained throughout adulthood. Moreover, cross-fostering studies report that these behavioral and molecular modifications are reversed when pups raised by low LG mothers are transferred to high LG mothers (Liu et al. 1997; Francis et al. 1999) within the first week of life.

8.4 The Glucocorticoid Receptor Gene

As the HPA programming by maternal behavior is modified by cross-fostering and temporally stable, researchers hypothesized that the long-term effects of maternal behavior and early-life environment variation on GR hippocampal expression could be due to epigenetic modifications. In rats, the GR gene is preceded by 10 noncoding exons and by 14 in humans (McCormick et al. 2000; Turner and Muller 2005). The expression of the noncoding exon 1₇ in rats and the human homologue 1_F have been shown to be specific to the hippocampus (Turner and Muller 2005). Each of the untranslated exon 1 variants has its own promoter and multiple transcription factor-binding sites, including NGFI-A (Meaney 2001), have been identified in GR promoter sequences (Turner et al. 2008, 2010). In offspring raised by low LG rat mothers, CpG methylation levels in the exon 1₇ promoter region are significantly increased at almost all CpGs compared to offspring raised by high LG mothers. More importantly, one CpG located in the 5' end of a NGFI-A-binding site is methylated in almost 100 % of offspring raised by low LG mothers whereas it is almost not methylated in offspring from high LG mothers (Weaver et al. 2004). Interestingly, follow-up studies in high and low LG rats showed that DNA methylation levels are also increased in the promoters of other noncoding first exons of the

GR gene which were associated with transcriptional changes (McGowan et al. 2011) suggesting that the whole GR locus may be poised for epigenetic regulation by environmental social factors.

These findings were recently translated to humans through studies investigating hippocampal tissue from individuals who died by suicide with and without a history of childhood adversity, as well as normal controls (McGowan et al. 2009; Labonte et al. 2012b). Notably, methylation levels in the exon 1_F promoter in abused suicide completers were significantly higher than among non-abused suicides and healthy controls. In addition, similarly to what was found in rats, a significant hypermethylation in an NGFI-A-binding site was found in abused suicide completers but not in the other groups. Through a series of cell functional assays, this epigenetic mark was shown to repress the binding of NGFI-A to its cognate DNA sequence and to decrease GR transcription (McGowan et al. 2009).

It is also interesting to note that these findings have been supported by other groups with different populations of individuals that suffered from early-life adversity. Indeed, higher levels of methylation in the promoter of GR 1_F have been reported in the infants of mothers reporting intimate partner violence during their pregnancy compared to those born from normal mothers (Radtke et al. 2011). Another study reported significant correlations between GR 1_F promoter methylation levels and parental loss, child maltreatment, and parental care (Tyrka et al. 2012). Furthermore, DNA methylation levels in GR 1_F promoter were shown to be positively correlated with childhood sexual abuse, its severity, and the number of maltreatment types in individuals with MDD, and with repetition of severe types of abuse in patients with bipolar disorders (Perroud et al. 2011). Altogether, this suggests that early-life adversity may induce specific long-lasting epigenetic alterations affecting gene expression.

In a different study assessing the expression of several GR exon 1 variants expressed in the limbic system of depressed suicide completers, GR1_F and GR1_C hippocampal expression were significantly decreased in depressed suicide completers (Alt et al. 2010). However, this was not associated with promoter hypermethylation although it should be noted that this study investigated methylation only in a limited region and promoter methylation levels reported were particularly low. On the other hand, NGFIA protein levels in the HPC were significantly decreased in depressed suicide completers suggesting that the decrease in GR expression found in suicide completers may be mediated by different molecular pathways depending on the presence or the absence of early-life adversity.

More recently, our group pushed further the investigation of early-life adversity consequences on the epigenetic regulation of GR in the HPC of abused suicide completers. Our data indicated that the expression of the noncoding exons 1_B, 1_C, and 1_H is significantly decreased in suicide completers with a history of childhood abuse compared to non-abused suicides and controls. The assessment of methylation levels in the promoter of GR1_C revealed methylation differences that are inversely correlated with GR1_C expression in accordance with our previous finding on 1_F variant. On the other hand, the GR1_H promoter showed site-specific

hypomethylation that was positively correlated with hGR1_H expression. In other words, lower levels of methylation were significantly correlated with lower expression, suggesting that active demethylation is also a functional mechanism that may be affected by early-life adversity. While this is a mechanism that has received less attention, more work is required in order to elucidate its potential implications in the context of early-life adversity.

In addition to DNA methylation, chromatin changes have also been associated with less frequent maternal stimulation. For instance, H3K9 acetylation, a marker of open euchromatin state (Kouzarides 2007), was found to be lower in the GR1₇ promoter in low LG raised rats. Pharmacological challenge with the histone deacetylase inhibitor (TSA) restored methylation levels, increased NGFI-A binding to the promoter, and reinstated H3K9 acetylation and GR hippocampal levels (Weaver et al. 2004). Treated rats were also less reactive to stressful conditions. By decreasing H3K9 acetylation, DNA access to the transcriptional machinery and DNA-binding proteins such as transcription factors and methylated DNA-binding proteins is reduced. Functionally, these results suggest that variation in the early-life environment in rats and early-life adversity in human induces a coordinated remodeling of epigenetic mechanisms involving DNA methylation and chromatin modifications in the multiple promoters of GR leading to important changes in GR expression, and consequent regulation of the HPA axis.

8.5 The Vasopressin (AVP) and Corticotropin-Releasing Factor Genes

Other components of the HPA axis have also been shown to be affected by early-life stress. For instance, early-life infant-maternal separation in mice, inducing stress-coping behavioral alterations in pups, has been shown to be associated with a long-lasting increase in corticosterone secretion and with an increased expression of POMC and AVP in the paraventricular nucleus (PVN) of the hypothalamus (Murgatroyd et al. 2009). The AVP gene in mice is composed of three coding exons and is oriented tail to tail with the *oxytocin* (*Oxt*) gene. Interestingly, the intergenic region between the AVP and the *Oxt* genes has been shown to include an enhancer modulating AVP expression (Gainer et al. 2001) and is itself composed of a CpG island (Murgatroyd et al. 2009).

Methylation at multiple sites within the AVP enhancer was shown to be decreased in the PVN of stressed mice 6 weeks, 3 months, and 1 year following the stress regimen (Murgatroyd et al. 2009). Consistent with the repressive role of DNA methylation on expression, this was associated with overexpression of AVP gene. The regulatory properties of this enhancer were defined by a deletion experiment. Deleting the first part of the enhancer partially reduced transcriptional activity, while removing the entire enhancer almost completely abolished the

gene's activity. Furthermore, methylation of the enhancer also significantly reduced transcriptional activity. Interestingly, AVP expression was also significantly increased in stressed mice at 10 days, although no methylation differences were observed in the AVP's enhancer. The AVP enhancer can putatively bind the methylated CpG-binding protein MeCP2. However, because of the repressive role of MeCP2 on transcription, one would expect the opposite tendency concerning AVP expression. MeCP2 has nevertheless been shown to be susceptible to inactivation by neuronal depolarization-induced phosphorylation, leading to its dissociation from putative targets (Chen et al. 2003; Zhou et al. 2006). Accordingly, higher neuronal activity-induced CaMKII immunoreactivity and phosphorylated MeCP2 levels have been reported in AVP-expressing neurons in the PVN of 10-day-old stress mice. Altogether, these results suggest that in young stressed mice, methylation patterns in AVP's enhancer allow the binding of MeCP2, which could then repress expression. However, since early-life stress also increases neuronal activity in AVP-expressing neurons, MeCP2 gets phosphorylated and inactivated. Consequently, the repressive effect of MeCP2 on AVP expression is abolished. On the other hand, methylation levels in AVP enhancer decrease with time. This may decrease MeCP2 binding and allow AVP to be expressed at higher levels. Overall, these results nicely suggest that alterations in DNA methylation found outside of the promoter might also be involved in physiological and behavioral modifications induced by environmental factors.

The regulation of a related peptide, CRF, has been recently shown to be also associated with epigenetic regulation by the social environment. Accordingly, CRF expression in the PVN of chronically socially defeated mice was found to be increased (Elliott et al. 2010). Interestingly, this effect was found only in animals susceptible to social stress and showing the normal subordinated behavior following chronic exposure to aggressive littermates as opposed to the resilient mice continuing to interact with their aggressor. This was associated with lower levels of methylation as reported by a reduced number of methylated clones in the susceptible group compared to control and resilient mice. A closer look at the methylation alteration induced by chronic social stress pointed to a single site of hypomethylation in the proximal promoter flanking the first exon and known to bind the cyclic adenosine monophosphate (cAMP) response element-binding protein (Aguilera et al. 2007). The importance of this site was further confirmed by luciferase assays showing that mutating a single base in the CRE-binding site substantially reduced the cAMP-induced CRF promoter activity (Elliott et al. 2010). These changes in methylation and expression were also accompanied by a significant decrease in the DNA methyltransferase 3b expression and by an increase in the expression of the demethylating candidate gadd45b. Interestingly, chronic treatment with the tricyclic antidepressant imipramine attenuated the changes in DNA methylation and expression levels induced by social stress (Elliott et al. 2010). Consequently, these findings in both the AVP and the CRF genes strongly support the involvement of active demethylation in the long-term effects of early-life adversity.

8.6 The Brain-Derived Neurotrophic Factor Gene

Neurotrophic factors are important candidate molecules to understand the development of psychopathology because of their role in neuronal survival and plasticity, as well as their expression in brain regions from the limbic system, where emotions and related behaviors are processed. For instance, it is hypothesized that their alteration could partly underlie changes in plasticity observed in the brains of suicides as well as the mood symptoms observed in depressive patients. While the major neurotrophic factors include nerve growth factor (NGF), neurotrophin 3 and 4 (NT3/4), fibroblast growth factor (FGF), transforming growth factor (TGF), and brain-derived neurotrophic factor (BDNF), the latter has received most of the attention in neurobiological research of psychiatric conditions such as depressive disorders and suicide. For instance, low serum and brain BDNF expression has been reported in patients with major depression (Brunoni et al. 2008; Dwivedi et al. 2003; Pandey et al. 2008) and these alterations were reversed by antidepressant treatment (Chen et al. 2001; Sen et al. 2008; Matrisciano et al. 2009). In mice, BDNF depletion induces depressive-like behaviors (Chan et al. 2006) while in rats, chronic stress and persistent pain reduce BDNF expression in the hippocampus (Gronli et al. 2006; Duric and McCarson 2005), and these effects are counteracted by antidepressant treatment (Duric and McCarson 2006; Rogoz et al. 2005; Xu et al. 2006).

BDNF epigenetic regulation has recently been investigated in mice and rat models of stress-induced depressive symptoms (Tsankova et al. 2006; Roth et al. 2009), as well as in a rat model of exposure to traumatic events (Roth et al. 2011). In both species, the BDNF gene contains nine 5' noncoding first exons with their own promoter, but coding for the same protein (Aid et al. 2007). The alternative splicing of these exons specifies the tissue in which BDNF is expressed (Aid et al. 2007). In both species, epigenetic processes involved in the transcriptional control of BDNF have been shown to be altered by stress. For instance, chronic social stress in mice decreases the expression of two specific BDNF transcripts (III and IV) in the hippocampus (Tsankova et al. 2006), while maternal maltreatment decreases prefrontal cortex (PFC) BDNF mRNA expression in rats (Roth et al. 2009). Although similar, these transcriptional alterations were shown to be induced by different epigenetic mechanisms. Indeed, chronic stress in mice raises H3K27 dimethylation levels in transcripts III and IV promoters (Tsankova et al. 2006), while site-specific hypermethylation is found in transcripts IV and IX promoters of maltreated rats (Roth et al. 2009). In the latter study, site-specific hypermethylation seems to follow a developmental pattern, with exon IX promoter hypermethylation occurring immediately after the maltreatment regimen, while promoter IV methylation increases gradually to reach significantly altered levels only at adulthood. Surprisingly, in one of these studies (Tsankova et al. 2006), no DNA methylation difference was found in association with histone modifications, while no histone modification was reported in association with DNA methylation alterations in the other study (Roth et al. 2009). These findings

illustrate that early-life or chronic stressors may alter different epigenetic mechanisms with common transcriptional consequences: the latter leading to the compaction of chromatin in its heterochromatic state, and the former, blocking the binding of transcription factors to DNA. On the other hand, these results may also highlight the heterogeneity of stress-induced epigenetic alterations between species.

Recently, the epigenetic regulation of BDNF has been shown to be altered in a rat model of PTSD symptoms (Roth et al. 2011). Given the findings discussed above, the authors focused on exon IV. Stressed rats showed increased DNA methylation in the dorsal dentate gyrus and in the CA3 regions. Contrary to expectation, a global hypomethylation was observed in the ventral CA1 region. These epigenetic alterations were accompanied by significant downregulation of BDNF exon IV expression in both the dorsal and ventral CA1 regions in the stressed rats relative to non-stressed rats. Interestingly, these alterations were restricted to the hippocampus since no alterations were found in the basolateral amygdala nor in the medial prefrontal cortex. These findings suggest that DNA methylation may be affected differently within the same structure depending on the function and the connections these regions have. Given that the expression of exon IV was decreased in the ventral CA1 without any significant changes in DNA methylation, these findings also suggest that, although DNA methylation may have an important role in the regulation of BDNF, other mechanisms are probably involved.

Pharmacological treatment with the tricyclic antidepressant imipramine was able to reverse the effect of chronic stress on BDNF transcription in mice (Tsankova et al. 2006). However, this reversal does not seem to be due to the reinstatement of altered histone modifications but rather due to alteration of an indirect pathway. Indeed, chronic but not acute imipramine treatment did not reinstate H3K27 basal dimethylation levels, but rather decreased HDAC5 levels in the hippocampus of chronically stressed mice leading to a global hyperacetylation in transcripts III and IV promoter regions. The importance of histone acetylation in the effect of antidepressant treatment has indeed been previously reported in animal models of stress-induced depression (Sun et al. 2013). Additionally, this hyperacetylation was associated with higher hippocampal levels of H3K4 dimethylation in the area of BDNF III and IV promoters with both modifications related to transcriptional activation. Consequently, these results suggest the existence of a compensatory mechanism in the reinstatement of basal BDNF levels by chronic imipramine treatment following chronic stress, and they emphasize the importance of chromatin hyperacetylation induced by antidepressant treatment.

Recently, the methylation state of BDNF was also assessed in postmortem brains from suicide completers (Keller et al. 2010). The human BDNF gene is also composed of 11 exons preceded by nine noncoding first exons regulating BDNF expression in different tissue (Pruunsild et al. 2007). In Keller et al. (2010) study, three different methods were used to quantify methylation levels in a region encompassing part of noncoding exon IV and its promoter in the Wernicke's area. Their results show that methylation in four CpGs located downstream of the promoter IV transcription initiation site were significantly increased in suicide

completers compared to controls. These differences were specific to the BDNF promoter since the investigation of genome-wide methylation in these subjects did not reveal any significant difference between groups. In addition, BDNF expression in subjects with high methylation levels was significantly lower than in subjects with low and medium methylation levels, supporting the repressive effects on transcription of methylation within the promoter.

Finally, following the studies in mice reported above (Tsankova et al. 2006), our group provided evidence suggesting that antidepressants promote open chromatin structure (i.e., lower H3K27me3 level) in the promoter of BDNF in the PFC (Chen et al. 2011). Follow-up studies in depressed patients also revealed higher BDNF IV noncoding exon expression in the blood of citalopram treatment responders compared to nonresponders (Lopez et al. 2012). Interestingly, H3K27me3 levels were inversely correlated with both BDNF IV expression levels and with the severity of symptoms.

8.7 The Ribosomal RNA Gene

Ribosomal RNA (rRNA) decodes the mRNA into amino acids. Hence, rRNA is a bottleneck structure for protein synthesis, allowing adequate cell function depending on the cell needs. The rRNA promoter is composed of two regulatory regions, namely, the upstream control element (UCE) and the core promoter that binds the upstream binding factor (UBF) (Haltiner et al. 1986; Learned et al. 1986; Ghoshal et al. 2004). The expression of rRNA genes has been shown to be epigenetically regulated both in mice (Santoro and Grummt 2001) and humans (Brown and Szyf 2007; Ghoshal et al. 2004). In mice, the recruitment of transcription repressors has been suggested to induce chromatin modifications leading to methylation of a single CpG found within UBF-binding sites in the UCE. This is thought to prevent UBF binding to its cognate sequence and to decrease rRNA expression (Santoro and Grummt 2001). In humans, despite the fact that the CpG density in both promoter regions differs from mice (Santoro and Grummt 2001; Ghoshal et al. 2004), rRNA expression has nevertheless been shown to be epigenetically regulated (Brown and Szyf 2007). Indeed, the active portion of the rRNA promoter associated with pol I has been shown to be completely unmethylated while the inactive portion is almost fully methylated (Brown and Szyf 2007).

The epigenetic control of rRNA gene expression has been shown to be dysregulated in the hippocampus of abused suicide completers (McGowan et al. 2008). Abused suicide completers exhibited smaller rRNA expression levels associated with increased methylation in 21 out of 26 CpGs found within the rRNA core promoter and UCE compared to controls. From a mechanistic point of view, these results suggest that methylation represses the interaction of the UBF with the core promoter sequence and consequently decreases both the recruitment of transcriptional cofactors and the transcriptional activity of the RNA pol. Interestingly, these alterations seem to be specific to the HPC since no group difference in the

rRNA methylation pattern was found in the cerebellum. In addition, these results did not reflect global methylation differences, as genome-wide methylation levels did not reveal any methylation difference between abused suicides and controls.

8.8 The Tropomyosin-Related Kinase B Receptor Gene

The transmembrane gene tropomyosin-related kinase B (TrkB) is the receptor for BDNF and has long been investigated in the neurobiology of mood and related disorders (Duman and Monteggia 2006; Kim et al. 2007; Dwivedi et al. 2003, 2009). Expression microarray studies have reported lower TrkB expression in the prefrontal cortex of depressed subjects (Aston et al. 2005; Nakatani et al. 2006) and antidepressant treatment has been shown to increase its expression in cultured astrocytes (Mercier et al. 2004).

The TrkB gene is found on chromosome 9 at locus q22.1 and has five splice variants. Splice variant T1 or TrkB-T1 is an astrocytic truncated form of TrkB lacking catalytic activity (Rose et al. 2003). Recently, analysis of the methylation pattern in the promoter of a subset of suicide completers with low levels of TrkB-T1 expression revealed two sites where methylation levels were higher in suicide completers compared to controls (Ernst et al. 2009b). The methylation pattern at those two sites was negatively correlated with the expression of TrkB-T1 in suicide completers, and this effect was specific to the prefrontal cortex, since no significant difference was found in the cerebellum. Such a pattern of expression and methylation is thought to increase predisposition to suicidal behaviors. In addition, suicide completers with low TrkB-T1 expression showed enrichment of H3K27 methylation in the TrkB promoter (Ernst et al. 2009a), suggesting that the astrocytic variant of TrkB may be under the control of epigenetic mechanisms involving histone modifications and DNA methylation. Interestingly, recent data showed that mice overexpressing the TrkB.T1 variant are more susceptible to chronic social stress than wild-type mice since the first group exhibits consistent social avoidance (Razzoli et al. 2011). Together, these data suggest that epigenetic changes in the TrkB.T1 promoter, inducing expression changes, could define the vulnerability to chronic social stress and possibly to early-life adverse experience.

8.9 The GABAergic System

The GABAergic system has been the focus of many research studies in postmortem brain samples of psychiatric patients, and particularly individuals with histories of depression (Klempman et al. 2009; Merali et al. 2004; Torrey et al. 2005), schizophrenia, or bipolar disorder, many of whom died by suicide (Guidotti et al. 2000; Heckers et al. 2002; Volk et al. 2000). For instance, reductions of reelin and glutamate decarboxylase 1 (GAD1) mRNA (Guidotti et al. 2000) and an increase

in DNMT1 expression (Veldic et al. 2004; Kundakovic et al. 2007) were previously reported in postmortem brains of schizophrenia and bipolar subjects who died by suicide. Promoter hypermethylation has been reported for both genes, consistent with the methylating role of DNMT1 (Grayson et al. 2005; Tamura et al. 2007).

More recently, the hippocampal expression of GAD1 has been shown to be affected by maternal care in rats (Zhang et al. 2010). Indeed, pups raised by mothers providing low LG have lower GAD1 hippocampal expression associated with promoter hypermethylation and lower levels of H3K9ac, compared to pups raised by high LG mothers. Interestingly, maternal LG is also associated with DNMT1 hippocampal levels. Functional assays revealed that the transcription factor NGFIA binds the GAD1 promoter in order to increase GAD1 expression. Consequently, these results suggest that, similar to the regulation of GR in rat hippocampus, GAD1 expression is modulated by maternal behavior via epigenetic mechanisms involving DNA methylation interfering with the binding of activating transcription factors and by chromatin modifications (Zhang et al. 2010).

These findings are in accordance with the study of Poulter et al. (2008) that examined the expression of DNA methyltransferases as well as the GABA_A receptor $\alpha 1$ subunit in the brain of suicide completers. Three hypermethylated CpG sites within the $\alpha 1$ subunit promoter were identified in the PFC of suicide completers and negatively correlated with DNMT3b protein expression. Besides DNMT3b, DNMT1 and DNMT3a levels have also been reported to be altered in the limbic system and brain stem of suicide completers. However, in this study there were no reports of early-life adversity, and thus, one cannot assume that these effects would be similar in abused suicide completers.

8.10 Other Epigenetic Alterations in Suicide Brains

In the light of the results discussed above, early-life adversity seems to modify epigenetic control of gene expression. These changes can take place through histone modifications and/or DNA methylation. Moreover, epigenetic changes correlate with behavioral modifications in animals and humans, thus strongly suggesting that epigenetics may act as an interface mediating the effect of environment on the genome.

Additional studies have focused on other functional systems which have been implicated in depression and suicide. Among these systems, the polyamine and the serotonergic systems are noteworthy.

Polyamines are ubiquitous aliphatic molecules involved in cellular functions including growth, division, and signaling cascades (Gilad and Gilad 2003; Minguet et al. 2008). The polyamines also play a major role in the regulation of stress (Rhee et al. 2007; Fiori and Turecki 2008), since they are dependent on the activation of the HPA axis and the subsequent increased concentrations of circulating glucocorticoids (Gilad and Gilad 2003). Furthermore, the emergence of the characteristic adult polyamine stress response correlates with the cessation of the

hyporesponsive period of the HPA axis system (Gilad et al. 1998). Previously, spermine synthase (SMS), spermidine/spermine N1-acetyltransferase (SAT1), and ornithine aminotransferase-like 1 (OATL1) expression have been shown to be altered in the limbic system of suicide completers with a history of depressive disorders (Sequeira et al. 2006, 2007). However, follow-up studies revealed that epigenetic alterations in the promoter region of genes involved in the polyamine synthesis do not account for these changes in expression. More recently, site-specific differential methylation has been found in the promoter of ARG2 and AMD1 (Gross et al. 2012) that was inversely correlated with expression in BA 44.

The serotonergic system is a neurotransmitter system of great importance in psychiatry and has been extensively investigated in depression and suicide. Lower concentrations, binding, neurotransmission, and reuptake of serotonin and its metabolites are risk markers for suicidality and major depression (Cronholm et al. 1977; Bhagwagar and Cowen 2008). Among the various serotonergic receptors, particular attention has been given to 5-HT_{2A} and its gene, as an important candidate in association studies of suicidal behavior (Du et al. 2001; Turecki et al. 1999). One of the variants most commonly investigated was the 102 C/T polymorphism, located in exon 1 (Du et al. 2000; de Luca et al. 2007). Methylation in the C-allele variant in this polymorphism has previously been associated with higher DNMT1 expression in the brain and leukocytes of healthy subjects (Polesskaya et al. 2006). Although methylation was reported as increased in leukocytes from suicide ideators, a nonsignificant hypomethylation was reported in the PFC of suicide completers carrying the C-allele (de Luca et al. 2009), suggesting that methylation levels may be different in individuals who committed suicide and those who are planning suicide. On the other hand, the functional significance of this hypermethylation in leukocytes remains to be explored, and since significance levels were not reached in brain tissue, further research is required.

8.11 Genome-Wide DNA Methylation Alteration by Early-Life Stress

Overall, environmental factors seem to target the epigenetic regulation of genes involved in key regulatory processes such as the HPA axis, neurotrophic factors, neurotransmission, polyamines, and protein synthesis. However, while a growing body of evidence supports the contribution of epigenetic factors translating the effects of ELA on the human genome, there is a real need for large-scale comprehensive studies assessing genome-wide epigenetic patterns in the context of different environmental factors. A few of these studies recently reported interesting findings suggesting that child abuse, while targeting critical genes, may also induce genome-wide reprogramming of DNA methylation patterns.

Our group recently assessed the impact of child abuse on genome-wide DNA methylation signatures in gene promoters (Labonte et al. 2012a). In this study, we compared hippocampal DNA methylation patterns between suicide completers with a severe history of child abuse (sexual and/or physical) and healthy controls. We identified hundreds of sites that were differentially methylated, both with increased and decreased methylation, in the hippocampus of severely abused suicide completers. It is interesting to note that DNA methylation levels in gene promoters were inversely correlated with gene expression at the genome-wide level, and differential methylation in abused suicides were enriched in genes involved in neuroplasticity, a finding consistent with the notion that abusive experiences during childhood lead to plastic changes in the brain as a response to these negative environmental stimuli. Similar observations have been made in suicide completers (Labonte et al. 2013) who present methylation changes that are enriched in genes related to learning and memory, and in peripheral samples from PTSD patients (Uddin et al. 2010). However, Uddin and colleagues did not restrict their analysis to promoters but rather measured DNA methylation levels at 14,000 CpGs across the genome. The analysis revealed an overrepresentation of differentially methylated CpGs in genes related to immune function. This may be translated into the development of different psychopathological processes. Furthermore, a previous study performed in the PFC of psychotic and bipolar patients reported differential methylation in numerous sites that were involved in glutamatergic and GABAergic neurotransmission, brain development, and response to stress (Mill et al. 2008). More importantly, these studies were conducted in different tissues (blood versus brain) and different brain regions (HPC versus PFC), which may account for the discrepancies between studies, as different tissues (Ladd-Acosta et al. 2007) and cell types (Deaton et al. 2011; Iwamoto et al. 2011) have been shown to exhibit specific DNA methylation signatures.

8.12 Conclusion

Examining together this extensive body of research, there is significant evidence suggesting that early-life adversity affects molecular mechanisms involved in the regulation of behavior. These effects involve alterations in DNA methylation and histone modifications, which are believed to induce behavioral aberrations during development or later in life by affecting genes involved in crucial neuronal processes. Studies performed in postmortem brains from suicide completers with a history of childhood abuse have highlighted several environmentally induced epigenetic alterations in the regulatory regions of genes involved in the response to stress (see Table 8.1 for a summary). Similarly, investigating the effect on animals of variation in early-life environment has revealed useful information for expanding our understanding of the molecular mechanisms involved in the effect of environmental stressors on the regulation of behavior (Table 8.2). Together, these

Table 8.1 Summary of published studies assessing epigenetic components in suicide

Studies	Brain regions	Genes	Findings
Labonté et al. (2013)	Hippocampus	Genome-wide	Genome-wide changes in DNA methylation levels in the suicide completers Inverse correlation between genome-wide DNA methylation and gene expression levels Enrichment of DNA methylation alterations in genes involved in learning and memory processes
Labonte et al. (2012a, b, 2013)	Hippocampus	GR	DNA methylation alterations in GR1B and 1C promoters in the suicide completers with a history of abuse ↓ Methylation in GR 1H promoter in the suicide completers with a history of abuse ↓ Expression of GR 1B, 1C, and 1H in the suicide completers with a history of abuse Inverse correlation between DNA methylation and expression levels of 1C and positive correlation for 1H
Lopez et al. (2012)	Blood	BDNF	↑ BDNF IV expression in responders to citalopram treatment compared to nonresponders Inverse correlation between H3K27me3 and BDNF IV expression levels
Chen et al. (2011)	PFC	BDNF	Inverse correlation between H3K27me3 levels and the severity of symptoms ↓ H3K27me3 levels in promoter IV in suicide completers
McGowan et al. (2009)	Hippocampus	GR	↑ Methylation in NGFI-A-binding site within GR promoter in the hippocampus of suicide completers with a history of abuse ↓ Expression of GR in the hippocampus of suicide completers with a history of abuse
Alt et al. (2010)	Amygdala Hippocampus Inf frontal gyrus Cingulate gyrus Nucleus accumbens	GR	<i>Amy</i> : ↓GRα protein, ↑ expression of 1 J, ↓ YY1 transcription factor <i>HPC</i> ; ↓ expression of GR1F and 1C, ↓ NGFIA transcription factor <i>IFC</i> ; ↓GRβ protein, ↓ YY1 and Sp1 transcription factors <i>CG</i> ; ↓GRα protein, ↑ expression of 1D, ↓ YY1, NGFIA and Sp1 transcription factors <i>Nac</i> ; ↓ expression of GR1B, ↓ NGFIA transcription factor, no methylation difference in promoters
McGowan et al. (2008)	Hippocampus	rRNA	Overall hypermethylation of rRNA promoter in the hippocampus of suicide completers with a history of abuse
Keller et al. (2010)	Wernicke's area	BDNF	↓ Expression of rRNA gene in the hippocampus of suicide completers with a history of abuse Hypermethylation at 4 CpGs within promoter/exon IV in suicide completers; negative correlation between BDNF promoter methylation levels and expression

(continued)

Table 8.1 (continued)

Studies	Brain regions	Genes	Findings
Ernst et al. (2009a, b)	Frontal cortex	TrkB-T1	↑ Methylation in two sites within the promoter of TrkB-T1 in the frontal cortex of suicide completers
Ernst et al. (2009a, b)	Frontal cortex	TrkB-T1	↓ Expression of TrkB-T1 in the frontal cortex of suicide completers ↑ H3K27 methylation in the frontal cortex of suicide completers Negative correlation between H3K27 methylation levels and TrkB-T1 expression in the frontal cortex of suicide completers
Poulter et al. (2008)	Fronto-polar cortex (FPC) Hippocampus Amygdala Brain stem	GABA _A α1 DNMT1 DNMT3a DNMT3b	FPC: ↓ expression of DNMT1 mRNA in suicide completers ↑ Expression of DNMT3b mRNA and protein levels in suicide completers ↑ Increased methylation at two sites in the promoter region of GABAA receptor subunit α1 in suicide completers <i>Limbic system</i> : ↓ expression of DNMT1 and DNMT3b mRNA levels in suicide completers <i>Brain stem</i> : ↓ expression of DNMT3b mRNA in suicide completers
Fiori and Turecki (2010)	PFC	SMOX, SMS	No effects of promoter's methylation in SMOX and SMS on expression levels
de Luca et al. (2009)	PFC	5-HT2A	↓ Methylation in the promoter region of 5-HT2A receptor associated with a C-allele (trend) in the suicide completers ↑ Methylation in the promoter region of 5-HT2A receptor associated with a C-allele in leukocytes of suicide attempters

Table 8.2 Summary of published studies assessing epigenetic components in animal models of stress-induced depressive symptoms

Studies	Animal model	Brain regions	Genes	Findings
Weaver et al. (2004)	Low/high licking and grooming	Hippocampus	GR	<i>In pups raised by LG mothers</i> Overall GR1 ₇ promoter hypermethylation ↓ H3K9 acetylation in GR 1 ₇ promoter ↓ binding of NGFIA in GR1 ₇ promoter
Tsankova et al. (2006)	Intruder test	Hippocampus	BDNF	<i>In stressed mice</i> ↓ expression of transcript III and IV ↑ H3K27dimethylation in transcript III and IV promoter No DNA methylation difference <i>Chronic treatment with imipramine</i> ↑ H3 acetylation and H3K4 dimethylation in transcript III and IV promoter ↓ HDAC5 in hippocampus of stress mice
Roth et al. (2009)	Stressed mothers	Prefrontal cortex	BDNF	<i>In maltreated rats</i> ↓ expression of BDNF transcript IX from childhood to adulthood ↓ expression of BDNF transcript IV at adulthood; overall hypermethylation in transcript IV promoter; transgenerational DNA methylation alterations in pups raised by abusive mothers
Roth et al. (2011)	PTSD animal model	Hippocampus Basolateral Amygdala Prefrontal cortex	BDNF	<i>In stressed rats</i> <i>Dorsal hippocampus</i> ↓ Expression of BDNF transcript IV ↑ Methylation in dentate gyrus ↑ Methylation in CA1 Ventral hippocampus ↓ Expression of BDNF transcript IV ↓ Methylation in CA3 <i>Basolateral amygdala</i> No change in methylation pattern in transcript IV <i>Prefrontal cortex</i> No change in methylation pattern in transcript IV
Murgatroyd et al. (2009)	Maternal deprivation	Hypothalamic paraventricular nucleus	AVP	<i>In stressed mice</i> ↑ expression of AVP Site-specific hypomethylation in AVP intergenic region phosphorylation of meCP2

(continued)

Table 8.2 (continued)

Studies	Animal model	Brain regions	Genes	Findings
Elliott et al. (2010)	Intruder test	PVN	CRF	<i>In stressed mice</i> ↑ Expression of CRF ↓ Methylation in CRF promoter (CRE) ↓ Expression of DNMT3b (acute), HDAC2 (long term) ↑ gadd45 (acute) <i>In resilient mice</i> No differential CRF expression No differential methylation <i>Chronic treatment with imipramine</i> Regulate behavior Regulate CRF expression Regulate CRF promoter methylation
Zhang et al. (2010)	Low/high licking and grooming	Hippocampus	GAD1	<i>In pups raised by low LG mothers</i> ↓ expression of GAD1 ↑ methylation within GAD1 promoter ↑ expression of DNMT1 ↓ H3K9 acetylation in GAD1 promoter

findings suggest that epigenetics may act as a mechanism whereby environmental factors act on the modulation of long-term behavioral responses. In individuals with particular predispositions toward psychiatric disorders, these alterations may help trigger the expression of the illness. From a therapeutic point of view, it is tempting to speculate on the clinical potential these findings may provide. In the future, they could potentially lead to the development of tools for the identification of individuals at risk, and therefore, the possibility of preventive intervention. However, there are major challenges in their potential implementation, not the least of which is access to target tissue in living subjects, modification of epigenetic profiles, and appropriate delivery of such interventions. Given that this is a relatively new area of research, the current knowledge is significantly limited. Integrating genome-wide approaches will provide a more comprehensive view on the complexity of the relationship between early-life adversity and the psychopathology of brain disorders. Furthermore, since these studies can provide information on the molecular nature of stress-induced psychopathologies, future work should assess whether similar alterations can be found in more accessible tissue.

References

- Agid O, Shapira B, Zislin J, Ritsner M, Hanin B, Murad H, Troudart T, Bloch M, Heresco-Levy U, Lerer B (1999) Environment and vulnerability to major psychiatric illness: a case control study of early parental loss in major depression, bipolar disorder and schizophrenia. *Mol Psychiatry* 4:163–172
- Aguilera G, Kiss A, Liu Y, Kamitakahara A (2007) Negative regulation of corticotropin releasing factor expression and limitation of stress response. *Stress* 10:153–161
- Aid T, Kazantseva A, Piirsoo M, Palm K, Timmusk T (2007) Mouse and rat BDNF gene structure and expression revisited. *J Neurosci Res* 85:525–535
- Alt SR, Turner JD, Klok MD, Meijer OC, Lakke EA, Derijk RH, Muller CP (2010) Differential expression of glucocorticoid receptor transcripts in major depressive disorder is not epigenetically programmed. *Psychoneuroendocrinology* 35:544–556
- Andersen SL, Tomada A, Vincow ES, Valente E, Polcari A, Teicher MH (2008) Preliminary evidence for sensitive periods in the effect of childhood sexual abuse on regional brain development. *J Neuropsychiatry Clin Neurosci* 20:292–301
- Aston C, Jiang L, Sokolov BP (2005) Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry* 10:309–322
- Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410:120–124
- Barrera LO, Li Z, Smith AD, Arden KC, Cavenee WK, Zhang MQ, Green RD, Ren B (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Res* 18:46–59
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837
- Bensley LS, van Eenwyk J, Spieker SJ, Schoder J (1999) Self-reported abuse history and adolescent problem behaviors. I. Antisocial and suicidal behaviors. *J Adolesc Health* 24:163–172

- Bhagwagar Z, Cowen PJ (2008) "It's not over when it's over": persistent neurobiological abnormalities in recovered depressed patients. *Psychol Med* 38:307–313
- Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127:1361–1373
- Bremner JD, Randall P, Vermetten E, Staib L, Bronen RA, Mazure C, Capelli S, McCarthy G, Innis RB, Charney DS (1997) Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biol Psychiatry* 41:23–32
- Bremner JD, Vythilingam M, Vermetten E, Southwick SM, McGlashan T, Nazeer A, Khan S, Vaccarino LV, Soufer R, Garg PK, Ng CK, Staib LH, Duncan JS, Charney DS (2003) MRI and PET study of deficits in hippocampal structure and function in women with childhood sexual abuse and posttraumatic stress disorder. *Am J Psychiatry* 160:924–932
- Briere J, Elliott DM (2003) Prevalence and psychological sequelae of self-reported childhood physical and sexual abuse in a general population sample of men and women. *Child Abuse Negl* 27:1205–1222
- Brown GW, Moran P (1994) Clinical and psychosocial origins of chronic depressive episodes. I: a community survey. *Br J Psychiatry* 165:447–456
- Brown SE, Szyf M (2007) Epigenetic programming of the rRNA promoter by MBD3. *Mol Cell Biol* 27:4938–4952
- Brunoni AR, Lopes M, Fregni F (2008) A systematic review and meta-analysis of clinical studies on major depression and BDNF levels: implications for the role of neuroplasticity in depression. *Int J Neuropsychopharmacol* 11:1169–1180
- Carpenter LL, Tyrka AR, McDougale CJ, Malison RT, Owens MJ, Nemeroff CB, Price LH (2004) Cerebrospinal fluid corticotropin-releasing factor and perceived early-life stress in depressed patients and healthy control subjects. *Neuropsychopharmacology* 29:777–784
- Carrion VG, Weems CF, Eliez S, Patwardhan A, Brown W, Ray RD, Reiss AL (2001) Attenuation of frontal asymmetry in pediatric posttraumatic stress disorder. *Biol Psychiatry* 50:943–951
- Chan JP, Unger TJ, Byrnes J, Rios M (2006) Examination of behavioral deficits triggered by targeting *Bdnf* in fetal or postnatal brains of mice. *Neuroscience* 142:49–58
- Chen B, Dowlathshahi D, Macqueen GM, Wang JF, Young LT (2001) Increased hippocampal BDNF immunoreactivity in subjects treated with antidepressant medication. *Biol Psychiatry* 50:260–265
- Chen ES, Ernst C, Turecki G (2011) The epigenetic effects of antidepressant treatment on human prefrontal cortex BDNF expression. *Int J Neuropsychopharmacol* 14:427–429
- Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* 302:885–889
- Children's Bureau (2010) Child maltreatment 2009. In: A. O. C., Youth and Families (ed) U.S. Department of health and human services. Government Printing Office, Washington, DC
- Clements A, Poux AN, Lo WS, Pillus L, Berger SL, Marmorstein R (2003) Structural basis for histone and phosphohistone binding by the GCN5 histone acetyltransferase. *Mol Cell* 12:461–473
- Cronholm B, Asberg M, Montgomery S, Schalling D (1977) Suicidal behaviour syndrome with low CSF 5-HIAA. *Br Med J* 1:776
- de Bellis MD, Keshavan MS, Frustaci K, Shifflett H, Iyengar S, Beers SR, Hall J (2002) Superior temporal gyrus volumes in maltreated children and adolescents with PTSD. *Biol Psychiatry* 51:544–552
- de Luca V, Likhodi O, Kennedy JL, Wong AH (2007) Differential expression and parent-of-origin effect of the 5-HT2A receptor gene C102T polymorphism: analysis of suicidality in schizophrenia and bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 144B:370–374
- de Luca V, Viggiano E, Dhoot R, Kennedy JL, Wong AH (2009) Methylation and QTDT analysis of the 5-HT2A receptor 102C allele: analysis of suicidality in major psychosis. *J Psychiatr Res* 43:532–537

- Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R, Bird A (2011) Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res* 21:1074–1086
- Dinwiddie S, Heath AC, Dunne MP, Bucholz KK, Madden PA, Slutske WS, Bierut LJ, Statham DB, Martin NG (2000) Early sexual abuse and lifetime psychopathology: a co-twin-control study. *Psychol Med* 30:41–52
- Driessen M, Herrmann J, Stahl K, Zwaan M, Meier S, Hill A, Osterheider M, Petersen D (2000) Magnetic resonance imaging volumes of the hippocampus and the amygdala in women with borderline personality disorder and early traumatization. *Arch Gen Psychiatry* 57:1115–1122
- Du L, Bakish D, Lapierre YD, Ravindran AV, Hrdina PD (2000) Association of polymorphism of serotonin 2A receptor gene with suicidal ideation in major depressive disorder. *Am J Med Genet* 96:56–60
- Du L, Faludi G, Palkovits M, Bakish D, Hrdina PD (2001) Serotonergic genes and suicidality. *Crisis* 22:54–60
- Duman RS, Monteggia LM (2006) A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 59:1116–1127
- Duric V, McCarson KE (2005) Hippocampal neurokinin-1 receptor and brain-derived neurotrophic factor gene expression is decreased in rat models of pain and stress. *Neuroscience* 133:999–1006
- Duric V, McCarson KE (2006) Effects of analgesic or antidepressant drugs on pain- or stress-evoked hippocampal and spinal neurokinin-1 receptor and brain-derived neurotrophic factor gene expression in the rat. *J Pharmacol Exp Ther* 319:1235–1243
- Dwivedi Y, Rizavi HS, Conley RR, Roberts RC, Tamminga CA, Pandey GN (2003) Altered gene expression of brain-derived neurotrophic factor and receptor tyrosine kinase B in postmortem brain of suicide subjects. *Arch Gen Psychiatry* 60:804–815
- Dwivedi Y, Rizavi HS, Zhang H, Mondal AC, Roberts RC, Conley RR, Pandey GN (2009) Neurotrophin receptor activation and expression in human postmortem brain: effect of suicide. *Biol Psychiatry* 65:319–328
- Elliott E, Ezra-Nevo G, Regev L, Neufeld-Cohen A, Chen A (2010) Resilience to social stress coincides with functional DNA methylation of the *Crf* gene in adult mice. *Nat Neurosci* 13:1351–1353
- Ernst C, Chen ES, Turecki G (2009a) Histone methylation and decreased expression of *TrkB.T1* in orbital frontal cortex of suicide completers. *Mol Psychiatry* 14:830–832
- Ernst C, Deleva V, Deng X, Sequeira A, Pomarenski A, Klempan T, Ernst N, Quirion R, Gratton A, Szyf M, Turecki G (2009b) Alternative splicing, methylation state, and expression profile of tropomyosin-related kinase B in the frontal cortex of suicide completers. *Arch Gen Psychiatry* 66:22–32
- Evans E, Hawton K, Rodham K (2005) Suicidal phenomena and abuse in adolescents: a review of epidemiological studies. *Child Abuse Negl* 29:45–58
- Fergusson DM, Horwood LJ, Lynskey MT (1996) Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *J Am Acad Child Adolesc Psychiatry* 35:1365–1374
- Fiori LM, Turecki G (2008) Implication of the polyamine system in mental disorders. *J Psychiatry Neurosci* 33:102–110
- Fiori LM, Turecki G, (2010) Genetic and epigenetic influences on expression of spermine synthase and spermine oxidase in suicide completers. *Int J Neuropsychopharmacol* 1–12
- Fischle W, Tseng BS, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H, Allis CD (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* 438:1116–1122
- Francis D, Diorio J, Liu D, Meaney MJ (1999) Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science* 286:1155–1158
- Gainer H, Fields RL, House SB (2001) Vasopressin gene expression: experimental models and strategies. *Exp Neurol* 171:190–199

- Ghoshal K, Majumder S, Datta J, Motiwala T, Bai S, Sharma SM, Frankel W, Jacob ST (2004) Role of human ribosomal RNA (rRNA) promoter methylation and of methyl-CpG-binding protein MBD2 in the suppression of rRNA gene expression. *J Biol Chem* 279:6783–6793
- Gilad GM, Gilad VH (2003) Overview of the brain polyamine-stress-response: regulation, development, and modulation by lithium and role in cell survival. *Cell Mol Neurobiol* 23:637–649
- Gilad GM, Gilad VH, Eliyavev Y, Rabey JM (1998) Developmental regulation of the brain polyamine-stress-response. *Int J Dev Neurosci* 16:271–278
- Gilbert R, Widom CS, Browne K, Fergusson D, Webb E, Janson S (2009) Burden and consequences of child maltreatment in high-income countries. *Lancet* 373:68–81
- Gladstone GL, Parker GB, Mitchell PB, Malhi GS, Wilhelm K, Austin MP (2004) Implications of childhood trauma for depressed women: an analysis of pathways from childhood sexual abuse to deliberate self-harm and revictimization. *Am J Psychiatry* 161:1417–1425
- Gorey KM, Leslie DR (1997) The prevalence of child sexual abuse: integrative review adjustment for potential response and measurement biases. *Child Abuse Negl* 21:391–398
- Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A, Costa E (2005) Reelin promoter hypermethylation in schizophrenia. *Proc Natl Acad Sci U S A* 102:9341–9346
- Gronli J, Bramham C, Murison R, Kanhema T, Fiske E, Bjorvatn B, Ursin R, Portas CM (2006) Chronic mild stress inhibits BDNF protein expression and CREB activation in the dentate gyrus but not in the hippocampus proper. *Pharmacol Biochem Behav* 85:842–849
- Gross JA, Fiori LM, Labonte B, Lopez JP, Turecki G (2012) Effects of promoter methylation on increased expression of polyamine biosynthetic genes in suicide. *J Psychiatr Res* 47 (4):513–519
- Guidotti A, Auta J, Davis JM, Di-Giorgi-gerevini V, Dwivedi Y, Grayson DR, Impagnatiello F, Pandey G, Pesold C, Sharma R, Uzunov D, Costa E (2000) Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. *Arch Gen Psychiatry* 57:1061–1069
- Haltiner MM, Smale ST, Tjian R (1986) Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. *Mol Cell Biol* 6:227–235
- Hata K, Okano M, Lei H, Li E (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129:1983–1993
- He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5:522–531
- Heckers S, Stone D, Walsh J, Shick J, Koul P, Benes FM (2002) Differential hippocampal expression of glutamic acid decarboxylase 65 and 67 messenger RNA in bipolar disorder and schizophrenia. *Arch Gen Psychiatry* 59:521–529
- Heim C, Mletzko T, Purselle D, Musselman DL, Nemeroff CB (2008a) The dexamethasone/corticotropin-releasing factor test in men with major depression: role of childhood trauma. *Biol Psychiatry* 63:398–405
- Heim C, Nemeroff CB (2001) The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol Psychiatry* 49:1023–1039
- Heim C, Newport DJ, Heit S, Graham YP, Wilcox M, Bonsall R, Miller AH, Nemeroff CB (2000) Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *JAMA* 284:592–597
- Heim C, Newport DJ, Mletzko T, Miller AH, Nemeroff CB (2008b) The link between childhood trauma and depression: insights from HPA axis studies in humans. *Psychoneuroendocrinology* 33:693–710
- Heim C, Shugart M, Craighead WE, Nemeroff CB (2010) Neurobiological and psychiatric consequences of child abuse and neglect. *Dev Psychobiol* 52:671–690
- Heim C, Young LJ, Newport DJ, Mletzko T, Miller AH, Nemeroff CB (2009) Lower CSF oxytocin concentrations in women with a history of childhood abuse. *Mol Psychiatry* 14:954–958
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466:1129–1133

- Iwamoto K, Bundo M, Ueda J, Oldham MC, Ukai W, Hashimoto E, Saito T, Geschwind DH, Kato T (2011) Neurons show distinctive DNA methylation profile and higher interindividual variations compared with non-neurons. *Genome Res* 21:688–696
- Jaffee SR, Moffitt TE, Caspi A, Fombonne E, Poulton R, Martin J (2002) Differences in early childhood risk factors for juvenile-onset and adult-onset depression. *Arch Gen Psychiatry* 59:215–222
- Kaplan MJ, Klinetob NA (2000) Childhood emotional trauma and chronic posttraumatic stress disorder in adult outpatients with treatment-resistant depression. *J Nerv Ment Dis* 188:596–601
- Keller S, Sarchiapone M, Zarrilli F, Videtic A, Ferraro A, Carli V, Sacchetti S, Lembo F, Angiolillo A, Jovanovic N, Pisanti F, Tomaiuolo R, Monticelli A, Balazic J, Roy A, Marusic A, Coccozza S, Fusco A, Bruni CB, Castaldo G, Chiariotti L (2010) Increased BDNF promoter methylation in the Wernicke area of suicide subjects. *Arch Gen Psychiatry* 67:258–267
- Kendler KS, Bulik CM, Silberg J, Hettema JM, Myers J, Prescott CA (2000) Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and cotwin control analysis. *Arch Gen Psychiatry* 57:953–959
- Kendler KS, Kuhn JW, Prescott CA (2004) Childhood sexual abuse, stressful life events and risk for major depression in women. *Psychol Med* 34:1475–1482
- Kim YK, Lee HP, Won SD, Park EY, Lee HY, Lee BH, Lee SW, Yoon D, Han C, Kim DJ, Choi SH (2007) Low plasma BDNF is associated with suicidal behavior in major depression. *Prog Neuropsychopharmacol Biol Psychiatry* 31:78–85
- Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Vermeulen M, Mann M, Bahler J, Green RD, Kouzarides T (2007) Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* 449:928–932
- Klempen TA, Sequeira A, Canetti L, Lalovic A, Ernst C, Ffrench-Mullen J, Turecki G (2009) Altered expression of genes involved in ATP biosynthesis and GABAergic neurotransmission in the ventral prefrontal cortex of suicides with and without major depression. *Mol Psychiatry* 14:175–189
- Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31:89–97
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128:693–705
- Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–930
- Kundakovic M, Chen Y, Costa E, Grayson DR (2007) DNA methyltransferase inhibitors coordinately induce expression of the human reelin and glutamic acid decarboxylase 67 genes. *Mol Pharmacol* 71:644–653
- Labonte B, Suderman M, Maussion G, Navaro L, Yerko V, Mahar I, Bureau A, Mechawar N, Szyf M, Meaney MJ, Turecki G (2012a) Genome-wide epigenetic regulation by early-life trauma. *Arch Gen Psychiatry* 69:722–731
- Labonte B, Yerko V, Gross J, Mechawar N, Meaney MJ, Szyf M, Turecki G (2012b) Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol Psychiatry* 72:41–48
- Labonte B, Suderman M, Maussion G, Lopez JP, Navaro-Sanchez L, Yerko V, Mechawar N, Szyf M, Meaney MJ, Turecki G. (2013) Genome-wide methylation changes in the brains of suicide completers. *Am J Psychiatry* 170:511–520
- Ladd-Acosta C, Pevsner J, Sabuncian S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP (2007) DNA methylation signatures within the human brain. *Am J Hum Genet* 81:1304–1315
- Lan F, Bayliss PE, Rinn JL, Whetstone JR, Wang JK, Chen S, Iwase S, Alpatov R, Issaeva I, Canaani E, Roberts TM, Chang HY, Shi Y (2007) A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 449:689–694
- Laplanche P, Diorio J, Meaney MJ (2002) Serotonin regulates hippocampal glucocorticoid receptor expression via a 5-HT7 receptor. *Brain Res Dev Brain Res* 139:199–203

- Learned RM, Learned TK, Haltiner MM, Tjian RT (1986) Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. *Cell* 45:847–857
- Liu D, Diorio J, Tannenbaum B, Caldji C, Francis D, Freedman A, Sharma S, Pearson D, Plotsky PM, Meaney MJ (1997) Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science* 277:1659–1662
- Lopez JP, Mamdani F, Labonte B, Beaulieu MM, Yang JP, Berlim MT, Ernst C, Turecki G (2012) Epigenetic regulation of BDNF expression according to antidepressant response. *Mol Psychiatry* 18(4):398–399
- Matrisciano F, Bonaccorso S, Ricciardi A, Scaccianoce S, Panaccione I, Wang L, Ruberto A, Tatarelli R, Nicoletti F, Girardi P, Shelton RC (2009) Changes in BDNF serum levels in patients with major depression disorder (MDD) after 6 months treatment with sertraline, escitalopram, or venlafaxine. *J Psychiatry Res* 43:247–254
- Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466:253–257
- McCauley J, Kern DE, Kolodner K, Dill L, Schroeder AF, Dechant HK, Ryden J, Derogatis LR, Bass EB (1997) Clinical characteristics of women with a history of childhood abuse: unhealed wounds. *JAMA* 277:1362–1368
- McCormick JA, Lyons V, Jacobson MD, Noble J, Diorio J, Nyirenda M, Weaver S, Ester W, Yau JL, Meaney MJ, Seckl JR, Chapman KE (2000) 5'-Heterogeneity of glucocorticoid receptor messenger RNA is tissue specific: differential regulation of variant transcripts by early-life events. *Mol Endocrinol* 14:506–517
- McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M, Turecki G, Meaney MJ (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* 12:342–348
- McGowan PO, Sasaki A, Huang TC, Unterberger A, Suderman M, Ernst C, Meaney MJ, Turecki G, Szyf M (2008) Promoter-wide hypermethylation of the ribosomal RNA gene promoter in the suicide brain. *PLoS One* 3:e2085
- McGowan PO, Suderman M, Sasaki A, Huang TC, Hallett M, Meaney MJ, Szyf M (2011) Broad epigenetic signature of maternal care in the brain of adult rats. *PLoS One* 6:e14739
- Meaney MJ (2001) Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu Rev Neurosci* 24:1161–1192
- Meaney MJ, Aitken DH, Sapolsky RM (1987) Thyroid hormones influence the development of hippocampal glucocorticoid receptors in the rat: a mechanism for the effects of postnatal handling on the development of the adrenocortical stress response. *Neuroendocrinology* 45:278–283
- Meaney MJ, Diorio J, Francis D, Weaver S, Yau J, Chapman K, Seckl JR (2000) Postnatal handling increases the expression of cAMP-inducible transcription factors in the rat hippocampus: the effects of thyroid hormones and serotonin. *J Neurosci* 20:3926–3935
- Mellen M, Ayata P, Dewell S, Kriacucionis S, Heintz N (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151:1417–1430
- Merali Z, Du L, Hrdina P, Palkovits M, Faludi G, Poulter MO, Anisman H (2004) Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA (A) receptor subunits in frontal cortical brain region. *J Neurosci* 24:1478–1485
- Mercier G, Lennon AM, Renouf B, Dessouroux A, Ramauge M, Courtin F, Pierre M (2004) MAP kinase activation by fluoxetine and its relation to gene expression in cultured rat astrocytes. *J Mol Neurosci* 24:207–216
- Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, Jia P, Assadzadeh A, Flanagan J, Schumacher A, Wang SC, Petronis A (2008) Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet* 82:696–711

- Minguet EG, Vera-Sirera F, Marina A, Carbonell J, Blazquez MA (2008) Evolutionary diversification in polyamine biosynthesis. *Mol Biol Evol* 25:2119–2128
- Mitchell JB, Iny LJ, Meaney MJ (1990) The role of serotonin in the development and environmental regulation of type II corticosteroid receptor binding in rat hippocampus. *Brain Res Dev Brain Res* 55:231–235
- Molnar BE, Berkman LF, Buka SL (2001) Psychopathology, childhood sexual abuse and other childhood adversities: relative links to subsequent suicidal behaviour in the US. *Psychol Med* 31:965–977
- Murgatroyd C, Patchev AV, Wu Y, Micale V, Bockmuhl Y, Fischer D, Holsboer F, Wotjak CT, Almeida OF, Spengler D (2009) Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat Neurosci* 12(12):1559–1566
- Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, Matsumoto I, Kato T, Osumi N, Higuchi T, Niwa S, Yoshikawa T (2006) Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. *Hum Mol Genet* 15:1949–1962
- Nelson CJ, Santos-Rosa H, Kouzarides T (2006) Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* 126:905–916
- Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, Firestein R, Cleary M, Jenuwein T, Herrera RE, Kouzarides T (2001) Rb targets histone H3 methylation and HP1 to promoters. *Nature* 412:561–565
- Okano M, Xie S, Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19:219–220
- Pandey GN, Ren X, Rizavi HS, Conley RR, Roberts RC, Dwivedi Y (2008) Brain-derived neurotrophic factor and tyrosine kinase B receptor signalling in post-mortem brain of teenage suicide victims. *Int J Neuropsychopharmacol* 11:1047–1061
- Pariante CM, Lightman SL (2008) The HPA axis in major depression: classical theories and new developments. *Trends Neurosci* 31:464–468
- Perroud N, Paoloni-Giacobino A, Prada P, Olie E, Salzmann A, Nicastro R, Guillaume S, Mouthon D, Stouder C, Dieben K, Huguelet P, Courtet P, Malafosse A (2011) Increased methylation of glucocorticoid receptor gene (NR3C1) in adults with a history of childhood maltreatment: a link with the severity and type of trauma. *Transl Psychiatry* 1:e59
- Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolzheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527
- Polesskaya OO, Aston C, Sokolov BP (2006) Allele C-specific methylation of the 5-HT2A receptor gene: evidence for correlation with its expression and expression of DNA methylase DNMT1. *J Neurosci Res* 83:362–373
- Poulter MO, Du L, Weaver IC, Palkovits M, Faludi G, Merali Z, Szyf M, Anisman H (2008) GABAA receptor promoter hypermethylation in suicide brain: implications for the involvement of epigenetic processes. *Biol Psychiatry* 64:645–652
- Pray-Grant MG, Daniel JA, Schieltz D, Yates JR 3rd, Grant PA (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* 433:434–438
- Pruunsild P, Kazantseva A, Aid T, Palm K, Timmusk T (2007) Dissecting the human BDNF locus: bidirectional transcription, complex splicing, and multiple promoters. *Genomics* 90:397–406
- Radtke KM, Ruf M, Gunter HM, Dohrmann K, Schauer M, Meyer A, Elbert T (2011) Transgenerational impact of intimate partner violence on methylation in the promoter of the glucocorticoid receptor. *Transl Psychiatry* 1:1–6
- Razzoli M, Domenici E, Carboni L, Rantamaki T, Lindholm J, Castren E, Arban R (2011) A role for BDNF/TrkB signaling in behavioral and physiological consequences of social defeat stress. *Genes Brain Behav* 10(4):424–433
- Rhee HJ, Kim EJ, Lee JK (2007) Physiological polyamines: simple primordial stress molecules. *J Cell Mol Med* 11:685–703
- Rogoz Z, Skuza G, Legutko B (2005) Repeated treatment with mirtazepine induces brain-derived neurotrophic factor gene expression in rats. *J Physiol Pharmacol* 56:661–671

- Rose CR, Blum R, Pichler B, Lepier A, Kafitz KW, Konnerth A (2003) Truncated TrkB-T1 mediates neurotrophin-evoked calcium signalling in glia cells. *Nature* 426:74–78
- Roth TL, Lubin FD, Funk AJ, Sweatt JD (2009) Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biol Psychiatry* 65:760–769
- Roth TL, Zoladz PR, Sweatt JD, Diamond DM (2011) Epigenetic modification of hippocampal Bdnf DNA in adult rats in an animal model of post-traumatic stress disorder. *J Psychiatr Res* 45 (7):919–926
- Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, Van de Vegte-Bolmer M, Selzer RR, Iniguez AL, Green RD, Sauerwein RW, Jensen ON, Stunnenberg HG (2009) Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 106 (24):9655–9660
- Sanders SL, Portoso M, Mata J, Bahler J, Allshire RC, Kouzarides T (2004) Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* 119:603–614
- Santa Mina EE, Gallop RM (1998) Childhood sexual and physical abuse and adult self-harm and suicidal behaviour: a literature review. *Can J Psychiatry* 43:793–800
- Santoro R, Grummt I (2001) Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Mol Cell* 8:719–725
- Santos-Rosa H, Schneider R, Bernstein BE, Karabetsou N, Morillon A, Weise C, Schreiber SL, Mellor J, Kouzarides T (2003) Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* 12:1325–1332
- Scher CD, Forde DR, McQuaid JR, Stein MB (2004) Prevalence and demographic correlates of childhood maltreatment in an adult community sample. *Child Abuse Negl* 28:167–180
- Sen S, Duman R, Sanacora G (2008) Serum brain-derived neurotrophic factor, depression, and antidepressant medications: meta-analyses and implications. *Biol Psychiatry* 64:527–532
- Sequeira A, Gwady FG, Ffrench-Mullen JM, Canetti L, Gingras Y, Casero RA Jr, Rouleau G, Benkelfat C, Turecki G (2006) Implication of SSAT by gene expression and genetic variation in suicide and major depression. *Arch Gen Psychiatry* 63:35–48
- Sequeira A, Klempan T, Canetti L, Ffrench-Mullen J, Benkelfat C, Rouleau GA, Turecki G (2007) Patterns of gene expression in the limbic system of suicides with and without major depression. *Mol Psychiatry* 12:640–655
- Smythe JW, Rowe WB, Meaney MJ (1994) Neonatal handling alters serotonin (5-HT) turnover and 5-HT2 receptor binding in selected brain regions: relationship to the handling effect on glucocorticoid receptor expression. *Brain Res Dev Brain Res* 80:183–189
- Stein MB, Koverola C, Hanna C, Torchia MG, McClarty B (1997) Hippocampal volume in women victimized by childhood sexual abuse. *Psychol Med* 27:951–959
- Sun H, Kennedy PJ, Nestler EJ (2013) Epigenetics of the depressed brain: role of histone acetylation and methylation. *Neuropsychopharmacology* 38:124–137
- Swigut T, Wysocka J (2007) H3K27 demethylases, at long last. *Cell* 131:29–32
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935
- Tamura Y, Kunugi H, Ohashi J, Hohjoh H (2007) Epigenetic aberration of the human REELIN gene in psychiatric disorders. *Mol Psychiatry* 12(519):593–600
- Tanskanen A, Hintikka J, Honkalampi K, Haatainen K, Koivumaa-Honkanen H, Viinamaki H (2004) Impact of multiple traumatic experiences on the persistence of depressive symptoms—a population-based study. *Nord J Psychiatry* 58:459–464
- Torrey EF, Barci BM, Webster MJ, Bartko JJ, Meador-Woodruff JH, Knable MB (2005) Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains. *Biol Psychiatry* 57:252–260
- Tsankova NM, Berton O, Renthal W, Kumar A, Neve RL, Nestler EJ (2006) Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat Neurosci* 9:519–525

- Tucker KL (2001) Methylated cytosine and the brain: a new base for neuroscience. *Neuron* 30:649–652
- Turecki G, Briere R, Dewar K, Antonetti T, Lesage AD, Seguin M, Chawky N, Vanier C, Alda M, Joobert R, Benkelfat C, Rouleau GA (1999) Prediction of level of serotonin 2A receptor binding by serotonin receptor 2A genetic variation in postmortem brain samples from subjects who did or did not commit suicide. *Am J Psychiatry* 156:1456–1458
- Turner JD, Alt SR, Cao L, Vernocchi S, Trifonova S, Battello N, Muller CP (2010) Transcriptional control of the glucocorticoid receptor: CpG islands, epigenetics and more. *Biochem Pharmacol* 80:1860–1868
- Turner JD, Muller CP (2005) Structure of the glucocorticoid receptor (NR3C1) gene 5' untranslated region: identification, and tissue distribution of multiple new human exon 1. *J Mol Endocrinol* 35:283–292
- Turner JD, Pelascini LP, Macedo JA, Muller CP (2008) Highly individual methylation patterns of alternative glucocorticoid receptor promoters suggest individualized epigenetic regulatory mechanisms. *Nucleic Acids Res* 36:7207–7218
- Tyrka AR, Price LH, Marsit C, Walters OC, Carpenter LL (2012) Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. *PLoS One* 7:e30148
- Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, De Los Santos R, Goldmann E, Galea S (2010) Epigenetic and immune function profiles associated with posttraumatic stress disorder. *Proc Natl Acad Sci U S A* 107:9470–9475
- Veldic M, Caruncho HJ, Liu WS, Davis J, Satta R, Grayson DR, Guidotti A, Costa E (2004) DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. *Proc Natl Acad Sci U S A* 101:348–353
- Volk DW, Austin MC, Pierri JN, Sampson AR, Lewis DA (2000) Decreased glutamic acid decarboxylase67 messenger RNA expression in a subset of prefrontal cortical gamma-aminobutyric acid neurons in subjects with schizophrenia. *Arch Gen Psychiatry* 57:237–245
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40:897–903
- Weaver IC, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M, Meaney MJ (2004) Epigenetic programming by maternal behavior. *Nat Neurosci* 7:847–854
- Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD (2005) WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121:859–872
- Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442:86–90
- Xiao T, Shibata Y, Rao B, Larabee RN, O'Rourke R, Buck MJ, Greenblatt JF, Krogan NJ, Lieb JD, Strahl BD (2007) The RNA polymerase II kinase Ctk1 regulates positioning of a 5' histone methylation boundary along genes. *Mol Cell Biol* 27:721–731
- Xu H, Chen Z, He J, Haimanot S, Li X, Dyck L, Li XM (2006) Synergetic effects of quetiapine and venlafaxine in preventing the chronic restraint stress-induced decrease in cell proliferation and BDNF expression in rat hippocampus. *Hippocampus* 16:551–559
- Zhang TY, Hellstrom IC, Bagot RC, Wen X, Diorio J, Meaney MJ (2010) Maternal care and DNA methylation of a glutamic acid decarboxylase 1 promoter in rat hippocampus. *J Neurosci* 30:13130–13137
- Zhou Z, Hong EJ, Cohen S, Zhao WN, Ho HY, Schmidt L, Chen WG, Lin Y, Savner E, Griffith EC, Hu L, Steen JA, Weitz CJ, Greenberg ME (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* 52:255–269
- Zlotnick C, Mattia J, Zimmerman M (2001) Clinical features of survivors of sexual abuse with major depression. *Child Abuse Negl* 25:357–367

Chapter 9

Interaction Between Genetics and Epigenetics in Cancer

Amanda Ewart Toland

Abstract Cancer is a disease caused by somatic mutations in key genes important in tumorigenesis. Historically, the focus of study has been on the role of mutations, including chromosomal rearrangements, in the pathogenesis of cancer. About 30 years ago, a link was made between methylation at promoter regions of tumor-suppressor genes and cancer development. Now we know that abnormal epigenetic regulation of genes through changes in methylation, chromatin remodeling, and expression of noncoding regulatory genes such as microRNAs all have been linked to cancer development, progression, and metastasis. Despite the classical definition of epigenetics as being independent of DNA sequence, DNA sequence variations, both somatic and germline, have been shown to influence specific epigenetic events. This chapter focuses on the relationship between DNA sequence and alterations in epigenetic patterning with a particular emphasis on human cancers.

9.1 Introduction

9.1.1 *Cancer Genetics and Genomics*

Cancer is a heterogeneous disease characterized by cells that exhibit abnormal growth and can spread and colonize other sites in the body. Cancer is described as a disease caused by mutations or changes to the DNA that result in features such as sustained proliferation, resistance to cell death, induction of apoptosis, ability to replicate indefinitely, evasion of suppressors of growth, and ability to invade and

A.E. Toland (✉)

Department of Internal Medicine, Division of Human Genetics, 998 Biomedical Research Tower, 460 W. 12th Ave., Columbus, OH 43210, USA
e-mail: amanda.toland@osumc.edu

metastasize (Hanahan and Weinberg 2011). Cancer-inducing mutations can be inherited from one's parents through the germline or can occur somatically during one's lifetime. Germline mutations exist in every cell of the body, whereas somatic mutations only occur in the original cell sustaining the mutation and all of its daughter cells. Tumors can contain tens to hundreds of different somatic mutations important in tumor development and progression. Germline mutations are not typically sufficient to drive tumorigenesis on their own, and hence additional somatic mutations are necessary.

There is a continuum of risk associated with DNA variations. Germline changes to the DNA that are typically rare and impart large magnitude of risk are considered to be mutations, also known as highly penetrant, pathogenic, or deleterious mutations. These mutations are associated with familial cancer syndromes such as Lynch syndrome (mutations in *MLH1*, *MSH2*, *MSH6*), hereditary breast and ovarian cancer (mutations in *BRCA1*, *BRCA2*), and von Hippel–Lindau syndrome (mutations in *VHL*). Germline changes to the DNA that are more common and exert low or moderate increases in risk are considered to be low-penetrance variants or moderate-risk alleles. Variations that occur at frequencies of above 1 % are classified as polymorphisms, although some low-penetrance alleles that impact cancer risk may be rarer than this. Individually, low-penetrance alleles do not increase an individual's risk sufficiently to be clinically useful. However, any one individual may carry hundreds of risk alleles which collectively, and in the context of specific environmental risk factors, can impart significant disease risk (Tenesa and Dunlop 2009). Low-penetrance variants may also modify the risk in individuals who carry a deleterious mutation and as such are called modifier alleles (Antoniou and Chenevix-Trench 2010).

9.1.2 Epigenetics

In addition to DNA sequence alterations, changes in gene regulation through epigenetic modifications can also initiate and promote tumorigenesis. In contrast to mutations which are defined by the DNA sequence itself, epimutations are changes which do not alter the DNA code, but modify the way in which genes are regulated or expressed. The most commonly described epigenetic events are CpG methylation of DNA base pairs and histone modifications that lead to chromatin remodeling. Additional epigenetic regulators of gene expression include microRNAs (miRNAs) and other small noncoding RNAs, paramutations, nuclear organization, and chromatin looping (Toland 2012). The epigenome, or all the epigenetic marks in a cell or an individual, is more fluid than DNA code. Epigenomes normally evolve over an individual's lifetime and are influenced by developmental stage, tissue type, environmental factors, and genetics (Gordon et al. 2012; Heyn et al. 2012; Meagher and Musser 2012). Like somatic mutations, epigenetic modifications can be passed from mother to daughter cells and can be environmentally induced. Unlike germline mutations that can be passed

from generation to generation and persist in the genome, most epigenetic alterations occurring during one's lifetime are reset during gametogenesis and early embryogenesis and are not transmitted to one's offspring (Morgan et al. 2005; Fleming et al. 2008; Migicovsky and Kovalchuk 2011). There is recent evidence, particularly in animal and plant models, that some epigenetic marks can persist or be reset and that soluble factors influencing gene expression and phenotype may be responsible (Migicovsky and Kovalchuk 2011; Cuzin and Rassoulzadegan 2010; Wagner et al. 2008; Grandjean et al. 2009).

9.1.2.1 Methylation

DNA nucleotides can be modified to induce alterations of gene regulation. The best characterized is DNA methylation which is the addition of a methyl group to a cytosine resulting in a 5-methylcytosine. In many cases, the cytosines are located 5' to a guanine (CpG). CpG islands are cytosine- and guanine-rich regions found in promoter regions of an estimated 70 % of genes and are also located near retrotransposable elements and repetitive DNA (Widschwendter and Jones 2002). The methylation process is characterized by the transfer of a methyl group from S-adenosylmethionine (SAM) by DNA methyltransferases (DNMTs) to the C5 position of a cytosine. Methylation of gene promoters is associated with silencing of transcription and hypomethylation with gene expression. Retrotransposons and repetitive DNA-containing regions are typically hypermethylated (Jintaridith and Mutirangura 2010). There is a strong correlation between DNA methylation of promoter regions with repressive histone modifications such as H3K27me3 that result in inactive chromatin states. Methylation of intronic regions has recently been recognized in actively transcribed genes and been postulated to have a role in pre-mRNA splicing (Shukla et al. 2011; Sati et al. 2012). A second type of DNA methylation, 5-hydroxymethylcytosine, has also been identified in gene promoters and intragenic regions, but the biological role of this modification is not fully understood (Jin et al. 2011; Ku et al. 2011).

9.1.2.2 Chromatin

Chromatin is a means of compacting DNA. It is also critical for determining accessibility of DNA to transcriptional machinery based on the chromatin state. Chromatin is characterized by 147 bp of DNA wrapped around a histone protein octamer consisting of four histone proteins, H2A, H2B, H3, and H4, each in duplicate (Kornberg and Lorch 1999). Transcriptionally active DNA occurs in open or non-condensed chromatin (euchromatin) whereas transcriptionally silenced DNA is found in highly compacted chromatin (heterochromatin). Chromatin states are determined by a large number of histone posttranslational modifications that include acetylation, methylation, ubiquitination, phosphorylation, and sumoylation (Fullgrave et al. 2011).

9.1.2.3 Chromatin Looping and Nuclear Positioning

Studies have indicated that the position of DNA in the nucleus, particularly in relationship to promoters and enhancers, may also be critical for gene regulation (Meister et al. 2010; Schoenfelder et al. 2010; Brickner et al. 2012; Burns and Wente 2012). This type of epigenetic regulation, also called position-effect variegation, was first described in *Drosophila* but has been found in several other organisms including humans (Guffei et al. 2010). In addition to position in the nucleus, the interaction between chromosomal regions has been found using new technologies such as chromosome conformation capture (3C) which cross-links DNA for downstream sequencing. Translocations, a hallmark of many cancer genomes, have been mapped to chromosomes sharing nuclear space suggesting the influence of the nuclear organization on specific translocations that occur in cancer (Zhang et al. 2012a). Single-nucleotide polymorphisms (SNPs) have been identified from genome-wide association studies that map to enhancer elements and disrupt normal regulation, possibly through interference with chromosome looping (Akhtar-Zaidi et al. 2012). Examples include variants that influence *SOX9* in prostate cancer and *c-MYC* in colorectal tumors (Wright et al. 2010; Zhang et al. 2012b).

9.1.2.4 Small Noncoding RNA

Another category of epigenetic regulation is noncoding RNAs. miRNAs are small noncoding RNAs of 18–23 nucleotides that regulate gene expression of specific genes by targeting mRNAs for degradation or inhibiting their translation. Additional noncoding RNAs, including long noncoding RNAs (lncRNAs), small nucleolar RNAs (sno-RNAs), piwi-interacting RNAs (piRNAs), promoter-associated small RNAs (pasRNAs), transcription initiation RNAs (tiRNAs), and endogenous small interfering RNAs (siRNAs) (Taft et al. 2010), are all associated with gene regulation via different mechanisms (Toland 2012). There is emerging evidence that some small noncoding RNAs may be transmitted to offspring through the sperm (Hamatani 2012).

9.2 Cancer and Epigenetics

Since 1983, when Feinberg and Vogelstein first described aberrant DNA methylation in cancers (Feinberg and Vogelstein 1983) countless studies have detailed the abnormal epigenetic patterning occurring in tumors. DNA hypomethylation of the genome and hypermethylation of critical tumor-suppressor genes are early events in several tumor types (Ehrlich 2006, 2009). Genomic instability, another feature of tumors, has been hypothesized to be caused in part by hypomethylation, possibly

through activation of viral genes or retrotransposable elements (Rodriguez et al. 2006; Daskalos et al. 2009). Whereas many of the epigenetic alterations that occur during tumorigenesis are thought to arise through environmental modifiers and somatic mutations in critical epigenetic regulators, some alterations are due to germline variations or mutations. There are several means in which alterations to the DNA sequence can influence epigenetic patterning and status of a cell including somatic mutations, germline mutations, germline variants, somatic copy number losses and amplifications, and copy number variants.

9.3 Changes in Germline DNA Leading to Aberrant Epigenetic Regulation in Tumorigenesis

9.3.1 Germline Mutations Leading to Cancer Syndromes

Germline mutations in genes responsible for normal epigenetic functioning have been linked to increased risk of developing cancer. Inherited mutations in *DICER*, a protein important in the processing of miRNA, are associated with familial pleuropulmonary blastoma (PPB)-predisposition syndrome (Hill et al. 2009) and are also found in multilocular cystic nephroma (Bahubeshi et al. 2010). Truncating somatic mutations in *TARBP2*, a component of the DICER complex, are mutated in colon and endometrial cell lines and primary tumors; these mutations result in aberrant miRNA processing (Melo et al. 2009).

In addition to mutations directly affecting the epigenetic machinery, inherited mutations have been associated with other cancer syndromes. The best studied hereditary cancer syndrome with links to inherited epigenetic alterations is Lynch syndrome, a hereditary colorectal cancer syndrome caused by germline mutations in the mismatch repair genes *MSH2*, *MLH1*, *MSH6*, and *PMS2*. There are multiple reports of epigenetic silencing of *MLH1* which persisted through multiple generations and led to Lynch syndrome cancers (Hitchins et al. 2005; Goel et al. 2011; Crepin et al. 2012). The original descriptions of this phenomenon postulated that the methylation was not tied to a particular *MLH1* mutation as no sequence changes were initially identified (Suter et al. 2004; Hitchins et al. 2011). Subsequently, one mechanism for *MLH1* epigenetic silencing in multiple Lynch syndrome families was identified as a single-nucleotide change in the 5'UTR of *MLH1* (Hitchins et al. 2011).

Another example of germline epigenetic alterations that leads to a high risk of cancer is *DAPK1*. *DAPK1*, a mediator of apoptosis, is epigenetically silenced via promoter methylation in familial chronic lymphocytic leukemia (Raval et al. 2007). A promoter mutation was found to interfere with *HOX7* binding leading to subsequent promoter methylation. Several studies have assessed hereditary breast and ovarian cancer families for constitutive promoter methylation suggestive of an epimutation similar to this. Initial studies using small number of *BRCA1/2* mutation-negative families did not find evidence of promoter methylation

(Chen et al. 2006), but larger studies suggest that a small percentage of these individuals may have inherited silencing of these genes (Snell et al. 2008; Wong et al. 2011; Hansmann et al. 2012). These studies suggest that other high-risk cancer syndromes may be due in part to epigenetic silencing although the etiology and the frequency of germline mutations that influence epigenetic patterning in hereditary cancer remain to be determined.

9.3.2 *Single-Nucleotide Polymorphisms and Allele-Specific Epigenetic Modifications*

Several studies have identified genes showing allele-specific expression; some of the allele-specific expression has been associated with allele-specific methylation, methylation that occurs preferentially on one allele compared to the other. Approximately 35,000 sites across the genome are estimated to undergo allele-specific methylation of which some may differentially affect expression (Schalkwyk et al. 2010). Methylation of key tumor-suppressor genes has been identified in the lungs of smokers. *MGMT* (O⁶-methylguanine-DNA methyltransferase) is a repair enzyme whose activity is thought to be protective against the mutagenic effects of carcinogens that induce O⁶-methylguanines. *MGMT* is silenced via promoter methylation in several cancers (Hegi et al. 2008). An SNP in the promoter region of *MGMT*, *rs16906252*, is associated with increased *MGMT* methylation in smokers and in colon cancer (Hawkins et al. 2009; Leng et al. 2011). Allele-specific methylation has been observed in genes with other roles in cancer. For example, allele-specific methylation of *ABCB1*, a gene whose expression is implicated in drug resistance, has been observed in breast cancer cell lines (Reed et al. 2010). Allele-specific expression can also be attributed to allele-specific difference in chromatin modifications. Allele-specific histone modifications have been identified in humans and chromatin states have been observed to vary between families (Kadota et al. 2007; Prendergast et al. 2012). These loci tend to correlate with regions demonstrating allele-specific expression. An association between allele-specific methylation and H3K27methylation has been identified (Statham et al. 2012). It is probable that future studies will identify genetic variants associated with allele-specific expression via modifications to the chromatin that will affect cancer risk or development.

9.3.3 *Copy Number Variations and Germline Mutations Leading to Aberrant Methylation*

SNPs are not the only type of genetic variation that can impact the epigenome. Copy number variations, polymorphic gains and losses of DNA, can also impact normal epigenetic patterning and cancer development. Germline mutations

resulting in loss or gain of large segments of DNA can also cause abnormal epigenetic regulation of genes important in cancer risk. A number of families with a clinical diagnosis of Lynch syndrome were identified that appeared to have familial methylation of the *MSH2* gene without any promoter or coding defects (Chan et al. 2006). One explanation of these results was that there was a heritable epimutation or allele-specific silencing of the *MSH2* gene in the absence of DNA alterations. Further analysis, however, revealed that deletions of the last two exons of the *EPCAM* gene, just upstream of *MSH2*, induced methylation of *MSH2* resulting in Lynch syndrome (Ligtenberg et al. 2009; Niessen et al. 2009). Another example of copy number deletions influencing colorectal cancer risk is duplication of *PTPRJ*. *PTPRJ* has been implicated as a candidate colorectal cancer susceptibility gene from mouse models (Ruivenkamp et al. 2002). A 170-kb intragenic duplication including the 5' end of the *PTPRJ* gene and resulting in promoter methylation was found in DNA of a family with early-onset CRC (Venkatachalam et al. 2010). As additional studies are conducted, it is likely that additional copy number variations or mutations will be identified that result in abnormal epigenetic patterning and increased cancer risk.

9.3.4 Single-Nucleotide Polymorphisms and microRNAs

miRNAs consist of 18–23 nucleotides that contain a seed region of 6–9 nucleotides. The seed region is believed to be critical for binding to target mRNAs. Thus, DNA variations that occur in the miRNA or the seed region of the potential target have the potential to disrupt the miRNA's ability to recognize and bind to its targets (Landi et al. 2012a). Conversely SNPs in the 3'UTR can interfere with binding of miRNAs. There are several examples of SNPs in miRNA-binding sites or miRNAs themselves that have been associated with an increase in cancer risk (Table 9.1) (Landi et al. 2012b). Three miRNAs, *miR-196a2*, *miR-499*, and *miR-146a*, contain SNPs in their pre-miRNA that are thought to affect processing and have been associated with increased risk of cancer. In 2008, *rs2910164* in pre-*miR-146a* was found to decrease the expression of mature *miR-146a* and increase the risk of developing papillary thyroid cancer (Jazdzewski et al. 2008). It was later associated with an increased risk of prostate, breast, ovarian, and gastric cancers among other cancer types (Shen et al. 2008; Xu et al. 2010; Zeng et al. 2010). One of the first genes to be identified with a 3'UTR variant affecting miRNA binding was *KRAS*. *KRAS* is an oncogene which is mutated at a high frequency in cancers of the colon, lung, and pancreas. SNP interferes with *let-7* binding to *KRAS* leading to an increase in *KRAS* expression. *Rs67164370* has been associated with an increase in risk of non-small-cell lung cancer and response to therapy for metastatic colorectal cancer (Chin et al. 2008; Zhang et al. 2011b). Another interesting gene containing a variant reported to affect miRNA binding is *SET8*. *SET8* is responsible for methylation of *TP53*, the “guardian of the genome” which regulates genomic stability. An SNP in the 3'UTR of *SET8* affects binding of *miR-502* resulting in altered *SET8* expression. Another SNP, *rs16917496*, is associated with younger age

Table 9.1 Germline variants affecting miRNAs or miRNA targets

Location of SNP	SNPs	Gene/miRNA	Cancer type	References
miRNA precursor	<i>rs2910164</i>	<i>miR-146a</i>	Prostate, thyroid, breast, ovarian, gastric	Jazdzewski et al. (2008), Shen et al. (2008), Xu et al. (2010), Zeng et al. (2010), Wang et al. (2012)
miRNA	<i>rs2292832</i>	<i>miR-149</i>	HNSCC	Liu et al. (2010)
miRNA precursor		<i>miR-499</i>	Breast	Hu et al. (2009)
miRNA	<i>rs11614913</i>	<i>miR-196a2</i>	Breast, lung	Hu et al. (2009), Tian et al. (2009)
miRNA	<i>rs16917496</i>	<i>miR-502/ SET8</i>	Breast	Song et al. (2009)
miRNA	<i>rs4919510</i>	<i>miR-608</i>	CRC recurrence/ death	Lin et al. (2012)
	<i>rs213210</i>	<i>miR-219-1</i>	CRC death	Lin et al. (2012)
3'UTR	<i>rs67164370</i>	<i>Kras/LET-7</i>	Lung cancer	Chin et al. (2008)
3'UTR	<i>rs1421</i>	<i>EPCAM/miR- 1183</i>	Breast cancer risk	Jiang et al. (2011)
3'UTR	<i>rs1044219</i>	<i>RY3/miR-367</i>	Breast cancer, survival	Zhang et al. (2011a)
3'UTR	<i>rs8126</i>	<i>TNFAI2</i>	HNSCC	Liu et al. (2011)
3'UTR		<i>IQGAP1/ miR-124</i>	Gastric cancer	Zheng et al. (2011)
Promoter	<i>rs4938723</i>	<i>miR-34b/c</i>	Hepatocellular carcinoma	Xu et al. (2011)
3'UTR	<i>rs709805</i>	<i>KIAA0182</i>	CRC	Landi et al. (2012b)
3'UTR	<i>rs354476</i>	<i>NUP210</i>	CRC	Landi et al. (2012b)
3'UTR	<i>rs799917</i>	<i>BRCA1/miR- 638</i>	Breast	Nicoloso et al. (2010)
3'UTR	<i>rs334248</i>	<i>TGFR1/miR- 187</i>	Breast	Nicoloso et al. (2010)

HNSCC head and neck squamous cell carcinoma, *CRC* colorectal cancer

of breast cancer onset (Song et al. 2009). In addition to SNPs in the miRNAs or their targets, SNPs in promoter regions of miRNAs, such as *miR-34b/c*, have also been associated with cancer. These studies demonstrate that genetic variants can have a significant impact on gene regulation via miRNAs and influence cancer risk.

9.4 Somatic Mutations and Epigenetic Patterning

9.4.1 Epigenetic Patterning of the Cancer Genome

Although many studies in the past have focused on specific loci or genes in studying epigenetic changes in cancer, it is now recognized that genome-wide changes to normal epigenetic patterning during cancer development are just as frequent.

There is also a link between epigenetics and genomic aberrations. Hypomethylation of the genome with specific hypermethylation of gene promoters is typical of several cancers (Ehrlich 2006). Hypomethylation of the genome is associated with an increase in DNA instability resulting in aneuploidy and the potential for additional epigenetic alterations (Ehrlich 2009; Rodriguez et al. 2006). In addition to methylation status, chromatin status, specifically H3K9me3 levels, is associated with an increase in the mutation rate (Schuster-Bockler and Lehner 2012; Bartolomei 2009). Thus, genomic changes that occur during tumorigenesis reflect complex interaction between the epigenome and genome.

9.4.2 Mutations in Epigenetic Regulators

Cancer is caused by a series of mutations characterized by loss-of-function mutations of tumor-suppressor genes such as *TP53* and activating mutations or amplification of oncogenes such as *KRAS*. Recently mutations in epigenetic regulatory genes, such as *DNMT3A*, *EZH2*, and *TET2*, have also been demonstrated to play a critical role in tumorigenesis (Chase and Cross 2011; McCabe et al. 2012; Perez et al. 2012; Shih et al. 2012). *EZH2* is an important component of the polycomb repressive complex 2 (PRC2) which is part of the machinery linking promoter methylation and chromatin silencing by catalyzing H3-K27 trimethylation. B-cell lymphomas with Y641F mutations in *EZH2* lead to increased levels of trimethylated histone H3K27, a marker of gene silencing (Yap et al. 2011). Other activating mutations of *EZH2* have also been described. *DNMT3A* is a critical enzyme important in the transfer of methyl groups to CpG dinucleotides leading to methylation. Whole-genome sequencing revealed that 22 % of cases of acute myeloid leukemia patients have mutations in *DNMT3A* (Ley et al. 2010). Mutations in this gene correlated with decreased methylation of multiple genes, although the exact implications of the induced hypomethylation are not yet clear. It is likely that mutations in additional genes that control epigenetics will be found as genome sequence of tumors for diagnostic and treatment decisions becomes more routine. From what we know thus far, it is clear that somatic mutations lead to changes in epigenetic patterning on a genomic level and vice versa.

9.4.3 Chromosomal Breakpoints and Methylation

Common types of mutations observed in cancer are chromosomal rearrangements and translocations. Many of the breakpoints defining these regions are common both between and within cancer types. Thus, there has been much speculation about the mechanism or the features of DNA sequence that characterize breakpoint hot spots. One study of over 100 breast tumors compared copy number data with methylation profiling and identified 93 of 217 common breakpoint loci as being

differentially methylated in tumors (Tang et al. 2012). The majority of these loci showed hypomethylation and about a third of them were located within 3 Mb of Alu SINE repetitive elements. In multiple myeloma, a correlation between the density of LINE-1 elements and common breakpoints was identified (Aoki et al. 2012). In addition, the relative degree of methylation of these LINE-1 elements was associated with poorer prognosis in this disease. Head and neck cancers positive for human papillomavirus infection also show a strong correlation with hypomethylation of LINE elements and loss of heterozygosity (Richards et al. 2009). These studies further support the link between hypomethylation and genomic instability in tumors.

9.4.4 Somatic Events Leading to microRNA Alterations

miRNA expression is frequently perturbed in cancers. Deletions of DNA containing miRNAs are one type of mutation leading to aberrant miRNA expression. Translocations, or juxtaposition of two different chromosomal regions, have also been associated with alterations in miRNA expression. In fact, the first miRNAs associated with cancer, *miR-15* and *miR-16*, were discovered because of a translocation at 13q14 in patients with chronic lymphocytic leukemia that led to a small deletion of 27 kb in a region devoid of genes (Calin et al. 2002). A translocation between chromosomes 15 and 17 is found in a subset of individuals with acute myeloid leukemia. These translocations are associated with elevated levels of *miR-127*, *miR-299*, *miR-370*, *miR-323*, and *miR-154* which all map to the same locus on human 14q32 (Dixon-McIver et al. 2008). The mechanism for this increase in expression is not known, but one hypothesis is that the translocation leads to a change in methylation or acetylation status of this locus.

9.5 Imprinting and Cancer

Imprinting is the differential epigenetic regulation of a gene which is dependent on the parent from which the gene or the allele is inherited. This results in a gene-dosage effect and is one example of methylation patterning that is reset early in development. Several genomic loci, encompassing approximately 100 genes, are known to be maternally or paternally imprinted and additional loci have recently been described as being imprinted or silenced on one allele in human and mouse models (Schuster-Bockler and Lehner 2012; Xie et al. 2012). Frequently, imprinted genes reside in clusters that contain an imprinting control region that shows parental specific methylation. Loss of imprinting (LOI) or abnormal imprinting of some of these regions has been linked to the development of cancer.

9.5.1 Imprinted Genes Important in Cancer Risk

One of the first loci in the genome identified as being imprinted is 11p15.5 (Zhang et al. 1993; Moulton et al. 1996). Wilms' tumor, a childhood tumor, is associated with LOI of the maternally expressed noncoding *H19* gene and the paternally expressed insulin-like growth factor two gene (*IGF2*) on chromosome 11p15.5 (Scott et al. 2008). LOI of *IGF2* is associated with several additional cancers including prostate, colon, and gastric (Cui et al. 2002, 2003; Bhusari et al. 2011; Zuo et al. 2011). There is some evidence that LOI of *IGF2* occurs in multiple tissues of an individual suggesting an early event in development (Cruz-Correa et al. 2004). Other developmental disorders associated with defects in imprinting or LOI include Prader–Willi syndrome (PWS) and McCune–Albright syndrome. PWS is caused by aberrant imprinting at the 15q11–13 locus. Cancers associated with PWS include myeloid leukemias (Davies et al. 2003). The risk of developing thyroid cancer, osteosarcoma, skin cancer, and neurofibromatosis are elevated in individuals with McCune–Albright syndrome who have imprinting alterations for *GNAS* on chromosome 20q13.2 (Chanson et al. 2007).

9.5.2 Imprinting of Noncoding Elements

Genes are not the only important elements in the genome that show imprinting. An estimated 7 % of known human miRNAs map to imprinting regions (Labialle and Cavaille 2011) and many of these show aberrant expression in cancers (Giradot et al. 2012). A pilot study of parent–child trios assessed the methylation status of LINE-1 elements in the blood and identified a high correlation of methylation levels between mother–daughter and father–daughter pairs, but only a weak correlation between mother–son and father–son pairs (Mirabello et al. 2010). In general, males showed a higher rate of LINE-1 methylation compared to females. Interestingly, in father–son pairs in which both the father and son had a diagnosis of testicular germ cell tumors, there was a correlation of methylation levels between the father and son. LINE-1 hypomethylation showed a marginal increase in the risk for testicular germ cell tumors. Mouse studies have also showed a link between inherited epigenetic modifiers in the risk of testicular germ cell tumors supporting the role of familial shared epigenetic patterning in susceptibility to this cancer (Lam et al. 2007).

9.5.3 Cancer Risk and Aberrant Imprinting

Although most described cases of aberrant imprinting and cancer are associated with syndromes, there are recent reports of parent of origin effects (POE) associated

with low-penetrance susceptibility alleles. The extent of POE is unknown as most genome-wide association studies are not able to determine which allele was inherited from which parent because only cases and controls are genotyped and not relatives of cases. DeCode Genetics genotyped a large percentage of the fairly homogenous Icelandic population which allowed reconstruction of common haplotypes and the ability to generate information on parent of origin. A few SNPs evaluated in the context of parental haplotypes show apparent allele-specific POE risks for cancer. One example is the C allele of *rs3817198* which maps to 11p15, a region known to be imprinted. The C allele is not associated with breast cancer when inherited from the mother, but shows a modest increase in risk when inherited from the father (Kong et al. 2009). Similarly, the T allele of *rs157935* on 7q32 is associated with basal cell carcinoma when inherited from the father, but not when inherited from the mother. These studies suggest that being able to determine the parent of origin of alleles may uncover low-penetrance risk alleles. Whereas the mechanism of the risk being dependent on parent of origin is assumed to be epigenetic, this has not been fully evaluated.

Sporadic cancers are not the only ones that exhibit POE. Germline mutations in the genes for subunits of enzymes in the respiratory chain complex II (succinate–ubiquinone oxidoreductase, succinate dehydrogenase) can lead to hereditary paraganglioma/pheochromocytoma (PGL/PCC) which are highly vascularized tumors of the paraganglia (Fishbein and Nathanson 2012). Five different types of hereditary PGL/PCC have been described. Whereas PGL types 3, 4, and 5 are inherited as autosomal dominant traits with mutations in *SDHC*, *SDHB*, and *SHDA*, PGL types 1 and 2 only occur when *SDHD* or *SDHAF2* mutations are inherited from the father. This suggests imprinting or silencing of the maternal allele. However, studies to examine methylation status and expression patterns of *SDHD* have not uncovered the mechanism for the POE (Müller 2011). In tumors loss of the maternal allele is required suggesting that there is some contribution of the maternal allele in suppressing tumorigenesis. A model for partial epigenetic silencing of the maternal allele has been proposed but the mechanism for this has not been established.

The penetrance of hereditary syndromes, such as Lynch syndrome, may also depend on the parent transmitting the mutation (van Vliet et al. 2011). A study of over 400 carriers of mutations in mismatch repair genes, *MLH1*, *MSH2*, and *MSH6*, showed that the average age of diagnosis for male carriers or obligate carriers with maternally derived mutations was 6 years earlier than those whose mutation was paternally derived. This observation also extended to higher CRC incidences in males whose mutations were inherited from their mothers. There was no difference seen in females whose mutations came from their mother versus their father. Other studies have found that the effect was seen in female offspring (Lindor et al. 2010). Whether these effects are epigenetic or due to in utero environmental exposures has not been determined.

The *retinoblastoma* or *RBI* gene is a classic tumor suppressor that is inactivated through mutations or DNA deletions in many tumors. Inherited mutations in the Rb tumor-suppressor gene lead to hereditary retinoblastoma, although variability in

penetrance in age of diagnosis has been noted in some families (Harbour 2001). In 2009, a CpG island in intron 2 of *RBI* was found to be preferentially methylated on the maternally inherited chromosome and unmethylated on the paternally inherited chromosome (Kanber et al. 2009). *RBI* shows alternative splicing and intron 2 acts as a promoter for one of these splice forms, but only on the chromosome inherited from the father. The alternative splice form is not expressed from the maternal chromosome. The clinical implications of this imprinting and splice-form variation are not known, but it may impact some of the differences observed within families with germline mutations.

9.6 Transgenerational Epigenetic Events

Transgenerational epigenetic events refer to the persistence of an epigenetic state through the germline to manifest in the next generation (Fleming et al. 2008). Abnormal epigenetic silencing or activation of genes was not thought to be inherited through the germline because during the formation of gametes and again post-fertilization, epigenetic marks including DNA methylation and histones are erased and reprogrammed (Migicovsky and Kovalchuk 2011). Germline DNA variants can confer differences in epigenetic regulation through effects on methylation and miRNA binding which can be inherited. There is also evidence from animal and plant models that epigenetic effects can escape reprogramming and persist through the germline in the absence of known DNA sequence alterations (Niksson et al. 2012). In *C. elegans* piRNAs can induce stable long-term epigenetic silencing of genes that can persist over many generations (Ashe et al. 2012). This has not been shown for humans, but is one potential mechanism for the observation of transgenerational epigenetic modifications that persist for multiple generations.

In humans, the evidence for multigenerational epigenetic inheritance not based on DNA sequence is less clear. To date, studies to test this hypothesis are difficult as any transgenerational epigenetic effects need to persist beyond the grandparent-to-grandchild transmission. This is because environmental triggers or exposure in a parent have been shown to induce epigenetic changes in the developing fetus and its gametes which could mimic a hereditary change (Skinner 2008). There are some reports of hereditary epigenetic silencing of *MLH1* leading to Lynch syndrome, a hereditary colorectal cancer syndrome, which persists through multiple generations (Gazzoli et al. 2002; Suter et al. 2004; Hitchins et al. 2005; Hitchins and Ward 2009; Goel et al. 2011; Crepin et al. 2012). One mechanism for this was identified as a variant in the 5'UTR of *MLH1* (Hitchins et al. 2011) suggesting that hereditary *MLH1* methylation in these families is due to differences in the DNA code. Epigenetic silencing of *KILLIN* in lymphocytes is associated with early-onset cancer and features of Cowden syndrome, but this observation has not been noted in multiple generations (Bennett et al. 2010). As the animal data for transgenerational effects on cancer risk is convincing, it is likely that researchers will obtain evidence for this phenomenon in humans as well.

9.7 Summary

Approximately 30 years ago, aberrant DNA methylation was first described in cancer. Since then changes in DNA methylation, chromatin modifications, expression of noncoding RNAs, and higher level organization of DNA in the nucleus have all been shown to be involved in promotion of key cellular and stromal changes that convert a cell to a tumor cell. Despite the classical definition of epigenetics as being independent of DNA sequence, DNA sequences influence specific epigenetic events. Germline and somatic mutations and sequence variations have been demonstrated to affect the normal epigenetic state through a variety of mechanisms and in doing so can lead to increases in cancer risk and tumor promotion (Fig. 9.1). As our understanding of less well-characterized epigenetic processes such as

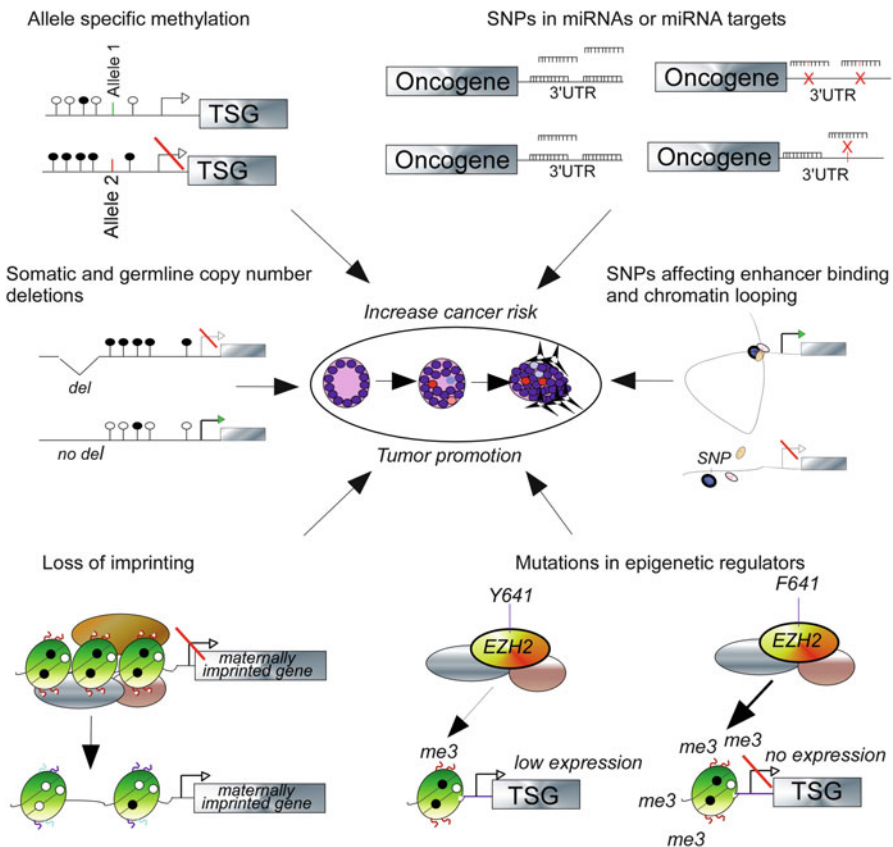


Fig. 9.1 Genetic influences on epigenetic patterning in cancer. Some of the different germline and somatic influences on epigenetic patterning that lead to an increase in cancer risk or which promote tumorigenesis are illustrated. *Filled circles*, methylated CpGs; *open circles*, unmethylated CpGs, *TSG*, tumor-suppressor gene; *del*, deletion

chromosome–chromosome interactions, transgenerational epigenetic effects, noncoding RNAs, and nuclear positioning becomes deeper, we will more fully understand the role of the genome and our individual genetic differences in shaping our epigenome and cancer development.

References

- Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakohova A, Bartels CF, Balasubramanian D et al (2012) Epigenomic enhancer profiling defines a signature of colon cancer. *Science* 336:736–739. doi:[10.1126/science.1217277](https://doi.org/10.1126/science.1217277)
- Antoniou AC, Chenevix-Trench G (2010) Common genetic variants and cancer risk in Mendelian cancer syndromes. *Curr Opin Genet Dev* 20:299–307. doi:[10.1016/j.gde.2010.03.010](https://doi.org/10.1016/j.gde.2010.03.010)
- Aoki Y, Nojima M, Suzuki H, Yasui H, Maruyama R, Yamamoto E et al (2012) Genomic vulnerability to LINE-1 hypomethylation is a potential determinant of the clinicogenetic features of multiple myeloma. *Genome Med* 4:101. doi:[10.1186/gm402](https://doi.org/10.1186/gm402)
- Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC et al (2012) piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 150:88–99. doi:[10.1016/j.cell.2012.06.018](https://doi.org/10.1016/j.cell.2012.06.018)
- Bahubeshi A, Bal N, Frio TR, Hamel N, Pouchet C, Yilmaz A et al (2010) Germline DICER1 mutations and familial cystic nephroma. *J Med Genet* 47:863–866. doi:[10.1136/jmg.2010.081216](https://doi.org/10.1136/jmg.2010.081216)
- Bartolomei MS (2009) Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev* 23:2314–2333. doi:[10.1101/gad.1841409](https://doi.org/10.1101/gad.1841409)
- Bennett KL, Mester J, Eng C (2010) Germline epigenetic regulation of KILLIN in Cowden and Cowden-like syndrome. *JAMA* 304:2724–2731. doi:[10.1001/jama.2010.1877](https://doi.org/10.1001/jama.2010.1877)
- Bhusari S, Yang B, Kueck J, Huang W, Jarrard DF (2011) Insulin-like growth factor-2 (IGF2) loss of imprinting marks a field defect within human prostates containing cancer. *Prostate* 71:1621–1630. doi:[10.1002/pros.21379](https://doi.org/10.1002/pros.21379)
- Brickner DG, Ahmed S, Meldi L, Thompson A, Light W, Young M et al (2012) Transcription factor binding to a DNA zip code controls interchromosomal clustering at the nuclear periphery. *Dev Cell* 22:1234–1246. doi:[10.1016/j.devcel.2012.03.012](https://doi.org/10.1016/j.devcel.2012.03.012)
- Burns LT, Wente SR (2012) Nuclear GPS for interchromosomal clustering. *Dev Cell* 22:1119–1120. doi:[10.1016/j.devcel.2012.05.019](https://doi.org/10.1016/j.devcel.2012.05.019)
- Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E et al (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 99:15524–15529. doi:[10.1073/pnas.242606799](https://doi.org/10.1073/pnas.242606799)
- Chan TL, Yuen ST, Kong CK, Chan YW, Chan ASY, Ng WF et al (2006) Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet* 38:1178–1183. doi:[10.1038/ng1866](https://doi.org/10.1038/ng1866)
- Chanson P, Salenave S, Orcel P (2007) McCune-Albright syndrome in adulthood. *Pediatr Endocrinol Rev* 4:452–462
- Chase A, Cross NC (2011) Aberrations of EZH2 in cancer. *Clin Cancer Res* 17:2613–2618. doi:[10.1158/1078-0432.CCR-10-2156](https://doi.org/10.1158/1078-0432.CCR-10-2156)
- Chen Y, Toland AE, McLennan J, Fridlyand J, Crawford B, Costello JF (2006) Lack of germ-line promoter methylation in BRCA1-negative families with familial breast cancer. *Genet Test* 10:281–284. doi:[10.1089/gte.2006.10.281](https://doi.org/10.1089/gte.2006.10.281)
- Chin LJ, Ratner E, Leng S, Zhai R, Nallur S, Babar I et al (2008) A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res* 68:8535–8540. doi:[10.1158/0008-5472.CAN-08-2129](https://doi.org/10.1158/0008-5472.CAN-08-2129)

- Crepin M, Dieu M-C, Lejeune S, Escande F, Boidin D, Porchet N et al (2012) Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. *Hum Mutat* 33:180–188. doi:[10.1002/humu.21617](https://doi.org/10.1002/humu.21617)
- Cruz-Correa M, Cui H, Giardiello FM, Powe NR, Hyland L, Robinson A et al (2004) Loss of imprinting of insulin growth factor II gene: a potential heritable biomarker for colon neoplasia predisposition. *Gastroenterology* 126:964–970. doi:[10.1053/j.gastro.2003.12.051](https://doi.org/10.1053/j.gastro.2003.12.051)
- Cui H, Onyango P, Brandenburg S, Wu Y, Hsieh CL, Feinberg AP (2002) Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Res* 62:6442–6446
- Cui H, Cruz-Correa M, Giardiello FM, Hutcheon DF, Kafonek DR, Brandenburg S et al (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 299:1753–1755. doi:[10.1126/science.1080902](https://doi.org/10.1126/science.1080902)
- Cuzin F, Rassoulzadegan M (2010) Non-Mendelian epigenetic heredity: gametic RNAs as epigenetic regulators and transgenerational signals. *Essays Biochem* 48:101–106
- Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R et al (2009) Hypomethylation of retrotransposable elements correlated with genomic instability in non-small cell lung cancer. *Int J Cancer* 124:81–87. doi:[10.1002/ijc.23849](https://doi.org/10.1002/ijc.23849)
- Davies HD, Leusink GL, McConnell A, Deyell M, Cassidy SB, Fick GH et al (2003) Myeloid leukemia in Prader-Willi syndrome. *J Pediatr* 142:174–178. doi:[10.1067/mpd.2003.81](https://doi.org/10.1067/mpd.2003.81)
- Dixon-McIver A, East P, Mein CA, Cazier J-B (2008) Distinctive patterns of microRNA expression associated with karyotype in acute myeloid leukaemia. *PLoS One* 3:e2141. doi:[10.1371/journal.pone.0002141](https://doi.org/10.1371/journal.pone.0002141)
- Ehrlich M (2006) Cancer-linked DNA hypomethylation and its relationship to hypermethylation. *Curr Top Microbiol Immunol* 310:251–274
- Ehrlich M (2009) DNA hypomethylation in cancer cells. *Epigenomics* 1:239–259. doi:[10.2271/epi.09.33](https://doi.org/10.2271/epi.09.33)
- Feinberg AP, Vogelstein B (1983) Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301:89–92
- Fishbein L, Nathanson KL (2012) Pheochromocytoma and paraganglioma: understanding the complexities of the genetic background. *Cancer Genet* 205:1–11. doi:[10.1016/j.cancergen.2012.01.009](https://doi.org/10.1016/j.cancergen.2012.01.009)
- Fleming JL, Huang TH, Toland AE (2008) The role of parental and grandparental epigenetic alterations in familial cancer risk. *Cancer Res* 68:9116–9121. doi:[10.1158/0008-5472.CAN-08-2814](https://doi.org/10.1158/0008-5472.CAN-08-2814)
- Fullgrabe J, Kavanagh E, Joseph B (2011) Histone onco-modifications. *Oncogene* 30:3391–3403. doi:[10.1038/onc.2011.121](https://doi.org/10.1038/onc.2011.121)
- Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD (2002) A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability high-tumor. *Cancer Res* 62:3925–3928
- Giradot M, Cavaille J, Feil R (2012) Small regulatory RNAs controlled by genomic imprinting and their contribution to human disease. *Epigenetics* 7:1341–1348. doi:[10.4161/epi.22884](https://doi.org/10.4161/epi.22884)
- Goel A, Nguyen TP, Leung HC, Nagasaka T, Rhee J, Hotchkiss E et al (2011) De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 128:869–878. doi:[10.1002/ijc.25422](https://doi.org/10.1002/ijc.25422)
- Gordon L, Joo JE, Powell JE, Ollikainen M, Novakovic B, Li X et al (2012) Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res* 22:1395–1406. doi:[10.1101/gr.136598.111](https://doi.org/10.1101/gr.136598.111)
- Grandjean V, Gounon P, Wagner N, Martin L, Wagner KD, Bernex F et al (2009) The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* 136:3647–3655. doi:[10.1242/dev.041061](https://doi.org/10.1242/dev.041061)

- Guffei A, Sarkar R, Klewes L, Righolt C, Knecht H, Mai S (2010) Dynamic chromosomal rearrangements in Hodgkin's lymphoma are due to ongoing three-dimensional nuclear remodeling and breakage-bridge-fusion cycles. *Haematologica* 95:2038–2046. doi:[10.3324/haematol.2010.030171](https://doi.org/10.3324/haematol.2010.030171)
- Hamatani T (2012) Human spermatozoa RNAs. *Fertil Steril* 97:275–281. doi:[10.1016/j.fertnstert.2011.12.035](https://doi.org/10.1016/j.fertnstert.2011.12.035)
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144:646–674. doi:[10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013)
- Hansmann T, Plushch G, Leubner M, Kroll P, Endt D, Gehrig A et al (2012) Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer. *Hum Mol Genet* 21:4669–4679. doi:[10.1093/hmg/dds308](https://doi.org/10.1093/hmg/dds308)
- Harbour JW (2001) Molecular basis of low-penetrance retinoblastoma. *Arch Ophthalmol* 119:1699–1704. doi:[10.1001/archophth.119.11.1699](https://doi.org/10.1001/archophth.119.11.1699)
- Hawkins NJ, Lee JH, Wong JJ, Kwok CT, Ward RL, Hitchins MP (2009) MGMT methylation is associated primarily with the germline C > T SNP (rs16906252) in colorectal cancer and normal colonic mucosa. *Mod Pathol* 22:1588–1599. doi:[10.1038/modpathol.2009.130](https://doi.org/10.1038/modpathol.2009.130)
- Hegi ME, Liu L, Herman JG, Stupp R, Wick W, Weller M et al (2008) Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. *J Clin Oncol* 26:4189–4199. doi:[10.1200/JCO.2007.11.5964](https://doi.org/10.1200/JCO.2007.11.5964)
- Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A et al (2012) Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* 109:10522–10527. doi:[10.1073/pnas.1120658109](https://doi.org/10.1073/pnas.1120658109)
- Hill DA, Ivanovich J, Priest JR, Gurnett CA, Dehner KP, Desruisseau D et al (2009) DICER1 mutations in familial pleuropulmonary blastoma. *Science* 325:965. doi:[10.1126/science.1174334](https://doi.org/10.1126/science.1174334)
- Hitchins MP, Ward RL (2009) Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J. Med Genet* 46:793–802. doi:[10.1136/jmg.2009.068122](https://doi.org/10.1136/jmg.2009.068122)
- Hitchins M, Williams R, Cheong K, Halani N, Lin VA, Packham D et al (2005) MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology* 129:1392–1399. doi:[10.1053/j.gastro.2005.09.003](https://doi.org/10.1053/j.gastro.2005.09.003)
- Hitchins MP, Rapkins RW, Kwok CT, Srivastava S, Wong JJ, Khachigian LM et al (2011) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. *Cancer Cell* 20:200–213. doi:[10.1016/j.ccr.2011.07.003](https://doi.org/10.1016/j.ccr.2011.07.003)
- Hu Z, Liang J, Wang Z, Tian T, Zhou X, Chen J et al (2009) Common genetic variants in pre-microRNAs were associated with increased risk of breast cancer in Chinese women. *Hum Mutat* 30:79–84. doi:[10.1002/humu.20837](https://doi.org/10.1002/humu.20837)
- Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A (2008) Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid cancer. *Proc Natl Acad Sci U S A* 105:7269–7274. doi:[10.1073/pnas.0802682105](https://doi.org/10.1073/pnas.0802682105)
- Jiang L, Zhang C, Li Y, Yu X, Zheng J, Zou P et al (2011) A non-synonymous polymorphism Thr115Met in the EpCAM gene is associated with an increased risk of breast cancer in Chinese population. *Breast Cancer Res Treat* 126:487–495. doi:[10.1007/s10549-010-1094-6](https://doi.org/10.1007/s10549-010-1094-6)
- Jin SG, Wu X, Li AX, Pfeiffer GP (2011) Genomic mapping of 5'hydroxymethylcytosine in the human brain. *Nucleic Acids Res* 39:5015–5024. doi:[10.1093/nar/gkr120](https://doi.org/10.1093/nar/gkr120)
- Jintaridith P, Mutirangura A (2010) Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol Genomics* 41:194–200. doi:[10.1152/physiolgenomics.00146.2009](https://doi.org/10.1152/physiolgenomics.00146.2009)
- Kadota M, Yang HH, Hu N, Wang C, Hu Y, Taylor PR et al (2007) Allele-specific chromatin immunoprecipitation studies show genetic influence on chromatin state in human genome. *PLoS Genet* 3:e81. doi:[10.1371/journal.pgen.0030081.g001](https://doi.org/10.1371/journal.pgen.0030081.g001)

- Kanber D, Berulava T, Ammerpohl O, Mitter D, Richter J, Siebert R et al (2009) The human retinoblastoma gene is imprinted. *PLoS Genet* 5:e1000790. doi:[10.1371/journal.pgen.1000790](https://doi.org/10.1371/journal.pgen.1000790)
- Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S et al (2009) Parental origin of sequence variants associated with complex diseases. *Nature* 462:868–874. doi:[10.1038/nature08625](https://doi.org/10.1038/nature08625)
- Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98:285–294. doi:[10.1016-S0092-8674\(00\)81958-3](https://doi.org/10.1016/S0092-8674(00)81958-3)
- Ku CS, Naidoo N, Wu M, Soong R (2011) Studying the epigenome using next generation sequencing. *J Med Genet* 48:721–730. doi:[10.1136/jmedgenet-2011-100242](https://doi.org/10.1136/jmedgenet-2011-100242)
- Labialle S, Cavaille J (2011) Do repeated arrays of regulatory small-RNA genes elicit genomic imprinting? Concurrent emergence of large clusters of small non-coding RNAs and genomic imprinting at four evolutionarily distinct eutherian chromosomal loci. *Bioessays* 33:565–573. doi:[10.1002/bies.201100032](https://doi.org/10.1002/bies.201100032)
- Lam MYJ, Heaney JD, Youngren KK, Kawasoe JH, Nadeau JH (2007) Trans-generational epistasis between Dnd1Ter and other modifier genes controls susceptibility to testicular germ cell tumors. *Hum Mol Genet* 16:2233–2240. doi:[10.1093/hmg/ddm175](https://doi.org/10.1093/hmg/ddm175)
- Landi D, Gemignani F, Landi S (2012a) Role of variations within microRNA-binding sites in cancer. *Mutagenesis* 27:205–210. doi:[10.1093/mutage/ger055](https://doi.org/10.1093/mutage/ger055)
- Landi D, Gemignani F, Pardini B, Naccarati A, Garritano S, Vodicka P et al (2012b) Identification of candidate genes carrying polymorphisms associated with the risk of colorectal cancer by analyzing the colorectal mutome and microRNAome. *Cancer* 118:4670–4680. doi:[10.1002/cncr.27435](https://doi.org/10.1002/cncr.27435)
- Leng S, Bernauer AM, Hong C, Do KC, Yingling CM, Flores KG et al (2011) The A/G allele of rs16906252 predicts for MGMT methylation and is selectively silenced in premalignant lesions from smokers and in lung adenocarcinomas. *Clin Cancer Res* 17:2014–2023. doi:[10.1158/1078-0432.CCR-10-3026](https://doi.org/10.1158/1078-0432.CCR-10-3026)
- Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE et al (2010) DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 363:2424–2433. doi:[10.1056/NEJMoa1005143](https://doi.org/10.1056/NEJMoa1005143)
- Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M et al (2009) Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 41:112–117. doi:[10.1038/ng.283](https://doi.org/10.1038/ng.283)
- Lin M, Gu J, Eng C, Ellis LM, Hildebrandt MA, Lin J et al (2012) Genetic polymorphisms in microRNA-related genes as predictors of clinical outcomes in colorectal adenocarcinoma patients. *Clin Cancer Res* 18:3982–3991. doi:[10.1158/1078-0432.CCR-11-2951](https://doi.org/10.1158/1078-0432.CCR-11-2951)
- Lindor NM, Rabe KG, Petersen GM, Chen H, Bapat B, Hopper J et al (2010) Parent of origin effects on age at colorectal cancer diagnosis. *Int J Cancer* 127:361–366. doi:[10.1002/ijc.25037](https://doi.org/10.1002/ijc.25037)
- Liu Z, Li G, Wei S, Niu J, El-Naggar AK, Sturgis EM et al (2010) Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck. *Cancer* 116:4753–4760. doi:[10.1002/cncr.25323](https://doi.org/10.1002/cncr.25323)
- Liu Z, Wei S, Ma H, Zhao M, Myers JN, Weber RS et al (2011) A functional variant at the miR-184 binding site in TNFAIP2 and risk of squamous cell carcinoma of the head and neck. *Carcinogenesis* 32:1668–1674. doi:[10.1093/carcin/bgr209](https://doi.org/10.1093/carcin/bgr209)
- McCabe MT, Graves AP, Ganji G, Diaz E, Halsey WS, Jiang Y et al (2012) Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* 109:2989–2994. doi:[10.1073/pnas.1116418109](https://doi.org/10.1073/pnas.1116418109)
- Meagher RB, Musser KJ (2012) The influence of DNA sequence on Epigenome-induced pathologies. *Epigenetics Chromatin* 5:11. doi:[10.1186/1756-8935-5-11](https://doi.org/10.1186/1756-8935-5-11)
- Meister P, Towbin BD, Pike BL, Ponti A, Gasser SM, Meister P, Towbin BD, Pike BL, Ponti A, Gasser SM (2010) The spatial dynamics of tissue-specific promoters during *C. elegans* development. *Genes Dev* 24:766–782. doi:[10.1101/gad.559610](https://doi.org/10.1101/gad.559610)

- Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, Calin GA et al (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet* 41:365–370. doi:[10.1038/ng.317](https://doi.org/10.1038/ng.317)
- Migicovsky Z, Kovalchuk I (2011) Epigenetic memory in mammals. *Front Genet* 2:28. doi:[10.3389/fgene.2011.00028](https://doi.org/10.3389/fgene.2011.00028)
- Mirabello L, Savage SA, Korde L, Gadalla SM, Greene MH (2010) LINE-1 methylation is inherited in familial testicular cancer kindreds. *BMC Med Genet* 11:77. doi:[10.1186/1471-2350-11-77](https://doi.org/10.1186/1471-2350-11-77)
- Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals. *Hum Mol Genet* 14:R47–R58. doi:[10.1093/hmg/ddi114](https://doi.org/10.1093/hmg/ddi114)
- Moulton T, Chung WY, Yuan L, Hensle T, Waber P, Nisen P, Tycko B (1996) Genomic imprinting and Wilms' tumor. *Med Pediatr Oncol* 27:476–483
- Müller U (2011) Pathological mechanisms and parent-of-origin effects in hereditary paraganglioma/pheochromocytoma (PGL/PCC). *Neurogenetics* 12:175–181. doi:[10.1007/s10048-011-0280-y](https://doi.org/10.1007/s10048-011-0280-y)
- Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M et al (2010) Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res* 70:2789–2798. doi:[10.1158/0008-5472.CAN-09-3541](https://doi.org/10.1158/0008-5472.CAN-09-3541)
- Niessen RC, Hofstra RM, Westers H, Ligtenberg MJ, Kooi K, Jager PO et al (2009) Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes Chromosomes Cancer* 48:737–744. doi:[10.1002/gcc.20678](https://doi.org/10.1002/gcc.20678)
- Niksson E, Larsen G, Manikkam M, Guerrero-Bosagno C, Savenkova MI, Skinner MK (2012) Environmentally induced epigenetic transgenerational inheritance of ovarian disease. *PLoS One* 7:e36129. doi:[10.1371/journal.pone.0036129](https://doi.org/10.1371/journal.pone.0036129)
- Perez C, Martinez-Calle N, Martin-Subero JJ, Segura V, Delabesse E, Fernandez-Mercado M et al (2012) TET2 mutations are associated with specific 5-methylcytosine and 5-hydroxymethylcytosine profiles in patients with chronic myelomonocytic leukemia. *PLoS One* 7:e31605. doi:[10.1371/journal.pone.0031605](https://doi.org/10.1371/journal.pone.0031605)
- Prendergast JG, Tong P, Hay DC, Farrington SM, Semple CA (2012) A genome-wide screen in human embryonic stem cells reveals novel sites of allele-specific histone modification associated with known disease loci. *Epigenetics Chromatin* 5:6. doi:[10.1186/1756-8935-5-6](https://doi.org/10.1186/1756-8935-5-6)
- Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen SS et al (2007) Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. *Cell* 129:879–890. doi:[10.1016/j.cell.2007.03.043](https://doi.org/10.1016/j.cell.2007.03.043)
- Reed K, Hembruff SL, Sprowl JA, Parissenti AM (2010) The temporal relationship between ABCB1 promoter hypomethylation, ABCB1 expression and acquisition of drug resistance. *Pharmacogenomics J* 10:489–504. doi:[10.1038/tpj.2010.1](https://doi.org/10.1038/tpj.2010.1)
- Richards KL, Zhang B, Baggerly KA, Colella S, Lang JC, Schuller DE, Krahe R (2009) Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability. *PLoS One* 4:e4941. doi:[10.1371/journal.pone.0004941](https://doi.org/10.1371/journal.pone.0004941)
- Rodriguez J, Frigola J, Vendrell E, Risques RA, Fraga MF, Morales C et al (2006) Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res* 66:8462–8468. doi:[10.1158/0008-5472.CAN-06-0293](https://doi.org/10.1158/0008-5472.CAN-06-0293)
- Ruivenkamp CA, van Wezel T, Zanon C, Stassen AP, Vlcek C, Csikos T et al (2002) Ptpn1 is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 31:295–300. doi:[10.1038/ng903](https://doi.org/10.1038/ng903)
- Sati S, Tanwar VS, Kumar KA, Patowary A, Jain V, Ghosh S et al (2012) High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. *PLoS One* 7:e31621. doi:[10.1371/journal.pone.0031621](https://doi.org/10.1371/journal.pone.0031621)
- Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN et al (2010) Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 86:196–212. doi:[10.1016/j.ajhg.2010.01.014](https://doi.org/10.1016/j.ajhg.2010.01.014)

- Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S et al (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* 42:53–61. doi:[10.1038/ng.496](https://doi.org/10.1038/ng.496)
- Schuster-Bockler B, Lehner B (2012) Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488:504–507. doi:[10.1038/nature11273](https://doi.org/10.1038/nature11273)
- Scott RH, Douglas J, Baskcomb L, Huxter N, Barker K, Hanks S et al (2008) Constitutional 11p15 abnormalities, including heritable imprinting center mutations, cause nonsyndromic Wilms tumor. *Nat Genet* 40:1329–1334. doi:[10.1038/ng.243](https://doi.org/10.1038/ng.243)
- Shen J, Ambrosone CB, DiCioccio RA, Odunsi K, Lele SB, Zhao H (2008) A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. *Carcinogenesis* 29:1963–1966. doi:[10.1093/carcin/bgn172](https://doi.org/10.1093/carcin/bgn172)
- Shih AH, Abdel-Wahab O, Patel JP, Levine RL (2012) The role of mutations in epigenetic regulators in myeloid malignancies. *Nat Rev Cancer* 12:599–612. doi:[10.1038/nrc3343](https://doi.org/10.1038/nrc3343)
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M et al (2011) CTCF-promoted DNA polymerase II pausing links DNA methylation to splicing. *Nature* 479:74–79. doi:[10.1038/nature10442](https://doi.org/10.1038/nature10442)
- Skinner MK (2008) What is an epigenetic transgenerational phenotype? F3 or F2. *Reprod Toxicol* 25:2–6. doi:[10.1016/j.reprotox.2007.09.001](https://doi.org/10.1016/j.reprotox.2007.09.001)
- Snell C, Krypyuk M, Wong EM, kConFab investigators, Loughrey MB, Dobrovic A (2008) BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. *Breast Cancer Res* 10:R12. doi:[10.1186/bcr1858](https://doi.org/10.1186/bcr1858)
- Song F, Zheng H, Liu B, Wei S, Dai H, Zhang L et al (2009) A miR-502 binding site single-nucleotide polymorphism in the 3'-untranslated region of the SET8 gene is associated with early age of breast cancer onset. *Clin Cancer Res* 15:6292–6300. doi:[10.1158/1078-0432.CCR-09-0826](https://doi.org/10.1158/1078-0432.CCR-09-0826)
- Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res* 22:1120–1127. doi:[10.1191/gr.132076.111](https://doi.org/10.1191/gr.132076.111)
- Suter CM, Martin DI, Ward (2004) Germline epimutation of MLH1 in individuals with multiple cancers. *Nat Genet* 36:497–501. doi:[10.1038/ng1342](https://doi.org/10.1038/ng1342)
- Taft RJ, Pang KC, Mercer TR, Dinger M, Mattick JS (2010) Non-coding RNAs: regulators of disease. *J Pathol* 220:126–139. doi:[10.1002/path.2638](https://doi.org/10.1002/path.2638)
- Tenesa A, Dunlop MG (2009) New insights in the aetiology of colorectal cancer from genome-wide association studies. *Nat Rev Genet* 10:353–358. doi:[10.1038/nrg2574](https://doi.org/10.1038/nrg2574)
- Tian T, Shu Y, Chen J, Hu Z, Xu L, Jin G et al (2009) A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. *Cancer Epidemiol Biomarkers Prev* 18:1183–1187
- Tang MH, Varadan V, Kamalakaran S, Zhang MQ, Dimitrova N, Hicks J (2012) Major chromosomal breakpoint intervals in breast cancer co-localized with differentially methylated regions. *Front Oncol* 2:197. doi:[10.3389/fonc.2012.00197](https://doi.org/10.3389/fonc.2012.00197)
- Toland AE (2012) Aberrant epigenetic regulation in breast cancer. In: Minarovits J, Niller HH (eds) *Patho-epigenetics of disease*. Springer, New York
- van Vliet CM, Dowty JG, van Vliet JL, Smith L, Mead LJ, Macrae FA et al (2011) Dependence of colorectal cancer risk on the parent-of-origin of mutations in DNA mismatch repair genes. *Hum Mutat* 32:207–212. doi:[10.1002/humu.21408](https://doi.org/10.1002/humu.21408)
- Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, Schackert HK, Görgens H, Hahn MM et al (2010) Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 139:2221–2224. doi:[10.1053/j.gastro.2010.08.063](https://doi.org/10.1053/j.gastro.2010.08.063)
- Wagner KD, Wagner N, Ghanbarian H, Grandjean V, Gounon P, Cuzin F et al (2008) RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev Cell* 14:962–969. doi:[10.1016/j.devcel.2008.03.009](https://doi.org/10.1016/j.devcel.2008.03.009)

- Wang M, Chu H, Li P, Yuan L, Fu G, Ma L et al (2012) Genetic variants in microRNAs predict bladder cancer risk and recurrence. *Cancer Res* 72:6173–6182. doi:[10.1158/0008-5472.CAN-12-0688](https://doi.org/10.1158/0008-5472.CAN-12-0688)
- Widschwendter M, Jones PA (2002) DNA methylation and breast carcinogenesis. *Oncogene* 21:5462–5482. doi:[10.1038/sj.onc.1205606](https://doi.org/10.1038/sj.onc.1205606)
- Wong EM, Southey MC, Fox SB, Brown MA, Dowty JG, Jenkins MA, Wong EM, Southey MC, Fox SB, Brown MA, Dowty JG, Jenkins MA et al (2011) Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. *Cancer Prev Res* 4:23–33. doi:[10.1158/1940-6207.CAPR-10-0212](https://doi.org/10.1158/1940-6207.CAPR-10-0212)
- Wright JB, Brown SJ, Cole MD (2010) Upregulation of C-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol* 30:1411–1420. doi:[10.1128/MCB.01384-89](https://doi.org/10.1128/MCB.01384-89)
- Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J et al (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148:816–831. doi:[10.1016/j.cell.2011.12.035](https://doi.org/10.1016/j.cell.2011.12.035)
- Xu B, Feng NH, Li PC, Tao J, Wu D, Zhang ZD et al (2010) A functional polymorphism in Pre-miR-146a gene is associated with prostate cancer risk and mature miR-146a expression in vivo. *Prostate* 70:467–472. doi:[10.1002/pros.21080](https://doi.org/10.1002/pros.21080)
- Xu Y, Liu L, Liu J, Zhang Y, Zhu J, Chen J et al (2011) A potentially functional polymorphism in the promoter region of miR-34b/c is associated with an increased risk for primary hepatocellular carcinoma. *Int J Cancer* 128:412–417. doi:[10.1002/ijc.25342](https://doi.org/10.1002/ijc.25342)
- Yap DB, Chu J, Berg T, Schapira M, Cheng SW, Moradian A et al (2011) Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* 117:2451–2459. doi:[10.1182/blood-2010-11-321208](https://doi.org/10.1182/blood-2010-11-321208)
- Zeng Y, Sun QM, Liu NN, Dong GH, Chen J, Yang L, Wang B (2010) Correlation between pre-miR-146a C/G polymorphism and gastric cancer risk in Chinese population. *World J Gastroenterol* 16:3578–3583. doi:[10.3748/wjg.v16.i28.3578](https://doi.org/10.3748/wjg.v16.i28.3578)
- Zhang Y, Shields T, Crenshaw T, Hao Y, Moulton T, Tycko B (1993) Imprinting of human H19: allele-specific CpG methylation, loss of the active allele in Wilms tumor, and potential for somatic allele switching. *Am J Hum Genet* 53:113–124. doi:[93/5301-0015](https://doi.org/10.2307/35301)
- Zhang L, Liu Y, Song F, Zheng H, Hu L, Lu H et al (2011a) Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification. *Proc Natl Acad Sci U S A* 108:13653–13658. doi:[10.1073/pnas.1103360108](https://doi.org/10.1073/pnas.1103360108)
- Zhang W, Winder T, Ning Y, Pohl A, Yang D, Kahn M et al (2011b) A let-7 microRNA-binding site polymorphism in 3'-untranslated region of KRAS gene predicts response in wild-type KRAS patients with metastatic colorectal cancer treated with cetuximab monotherapy. *Ann Oncol* 22:104–109. doi:[10.1093/annonc/mdq315](https://doi.org/10.1093/annonc/mdq315)
- Zhang X, Cowper-Sal Lari R, Bailey SD, Moore JH, Lupien M (2012a) Integrative functional genomics identifies an enhancer looping to the SOX9 gene disrupted by the 17q24.3 prostate cancer risk locus. *Genome Res* 22:1437–1446. doi:[10.1101/gr.135665.111](https://doi.org/10.1101/gr.135665.111)
- Zhang Y, McCord RP, Ho Y-J, Lajoie BR, Hildebrand DG, Simon AC et al (2012b) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148:908–921. doi:[10.1016/j.cell.2012.02.002](https://doi.org/10.1016/j.cell.2012.02.002)
- Zheng H, Song F, Zhang L, Yang D, Ji P, Wang Y et al (2011) Genetic variants at the miR-124 binding site on the cytoskeleton-organizing IQGAP1 gene confer differential predisposition to breast cancer. *Int J Oncol* 38:1153–1161. doi:[10.3892/ijo.2011.940](https://doi.org/10.3892/ijo.2011.940)
- Zuo QS, Yan R, Feng DX, Zhao R, Chen C, Jiang YM et al (2011) Loss of imprinting and abnormal expression of the insulin-like growth factor 2 gene in gastric cancer. *Mol Carcinog* 50:390–396. doi:[10.1002/mc.20731](https://doi.org/10.1002/mc.20731)

Part III
Impact of Epigenetics on Complex
Trait Genetics and Analysis

Chapter 10

Epigenetic Variation, Phenotypic Heritability, and Evolution

Robert E. Furrow, Freddy B. Christiansen, and Marcus W. Feldman

Abstract Familial aggregation of complex diseases may have many causes in addition to and apart from genetic predisposition due to common ancestry. For example, exposure to an environment that induces susceptibility to a disease may produce similar familial aggregations when the environment is shared by family members. In general, according to the principles of (Johannsen 1903), the emergence of a disease phenotype is the result of the combined effects of the genotype of the individual and the environment that it experiences during development. The heritability of a disease is a measure of familial aggregation in terms of the covariances among family members relative to the variance in disease state in the general population. Thus heritability expresses the within-family resemblance, observed by Darwin and inferred by him to reflect inheritance, which formed the core of his (Darwin 1859) theory of evolution. Darwin's inspiration originated from the practical use of family resemblance in animal breeding. Animal breeders have long known that a major obstacle to progress in genetic improvement is the interaction between familial aggregation of environments and the effects of similar genetics within families. The potential importance of this interaction, recognized in classical studies of the genetic epidemiology of complex diseases and other quantitative characters, has reemerged in studies of the effects of epigenetic modifications, their variation, and their transmission between generations.

R.E. Furrow (✉) • M.W. Feldman
Department of Biology, Stanford University, Stanford, CA 94305, USA
e-mail: rfurrow@stanford.edu

F.B. Christiansen
Department of Bioscience and Bioinformatics Research Center, University of Aarhus,
DK-8000 Aarhus C, Denmark

10.1 Introduction

Familial aggregation of complex diseases may have many causes in addition to and apart from genetic predisposition due to common ancestry. For example, exposure to an environment that induces susceptibility to a disease may produce similar familial aggregations when the environment is shared by family members. In general, according to the principles of Johannesssen (1903), the emergence of a disease phenotype is the result of the combined effects of the genotype of the individual and the environment that it experiences during development. The heritability of a disease is a measure of familial aggregation in terms of the covariances among family members relative to the variance in disease state in the general population. Thus heritability expresses the within-family resemblance, observed by Darwin and inferred by him to reflect inheritance, which formed the core of his (Darwin 1859) theory of evolution. Darwin's inspiration originated from the practical use of family resemblance in animal breeding. Animal breeders have long known that a major obstacle to progress in genetic improvement is the interaction between familial aggregation of environments and the effects of similar genetics within families. The potential importance of this interaction, recognized in classical studies of the genetic epidemiology of complex diseases and other quantitative characters, has reemerged in studies of the effects of epigenetic modifications, their variation, and their transmission between generations.

Epigenetic modification patterns are known to exhibit dependence on the environment to which an organism's cells have been exposed (e.g., Skinner 2011). During ontogenesis epigenetic status evolves, and the status of gene action within the cells of an organism is a function of its epigenetic status. In addition, the external environment of the organism may influence the presence or absence of specific epigenetic signals in one or more tissues (Carone et al. 2010; Heijmans et al. 2008; Kucharski et al. 2008; McGowan et al. 2009; Ng et al. 2010; Sandovici et al. 2011; Tobi et al. 2009; Verhoeven et al. 2010; Waterland and Jirtle 2003; Weaver et al. 2004). As with the effects of radiation exposure, the full epigenetic impact of an exposure to a specific environment may take generations to emerge. A pregnant woman for example carries three generations of genomes (Fig. 10.1), as a result of which there may be apparent inheritance of the effects of environmental exposures over several generations. Skinner (2011) described such an epigenetic effect due to exposure to the fungicide vinclozolin. During gestation an exposed mother can transmit the causative agent to her fetus, resulting in epigenetic modifications. It also influences the germline contributing to her future grandchildren. The exposures of the mother and fetus are truly simultaneous, and so are the exposures of the eggs of a female fetus and the germline stem cells of a male fetus. In this case, transmission of induced epigenetic modifications may reveal itself three generations after the exposure ceases.

Familial aggregation of epigenetic modifications may thus emerge from two different sources, namely through the transmission of epigenetic marks and because of shared familial environments experienced by individuals in the population.

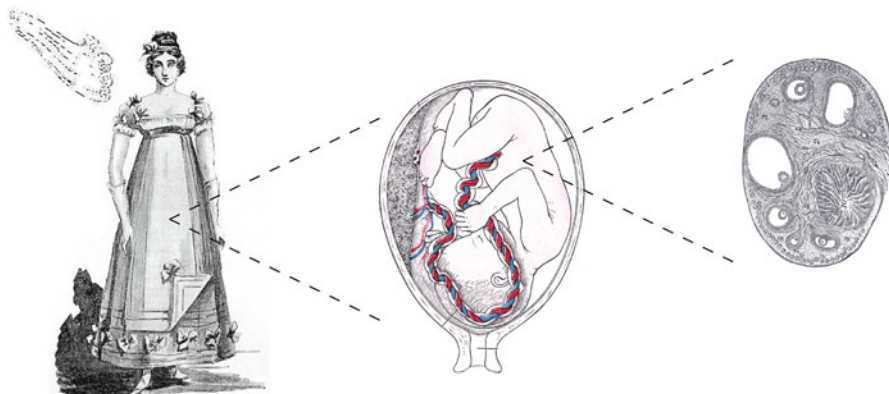


Fig. 10.1 Three generations of cells within one individual. Within a pregnant woman are the cells of her fetus, in addition to the germ cells of the fetus itself

These are not mutually exclusive. Transmission of epigenetic modifications in terms of methylation status of CpG sites is well understood in somatic cells. Similarities of epigenetic modifications among parents and their progeny are known, but seem at variance with the widespread removal of methylation from CpG sites in the early mammalian embryo. Nevertheless familial aggregations of methylation patterns have been observed, and recently incomplete resetting, transposon effects, and small RNA molecules have been observed to play a role in the transgenerational memory of epigenetic states (reviewed in Daxinger and Whitelaw 2012).

Future studies of the biological bases for a phenotype may be able to include data from many levels between genetics and a specific phenotype, including gene expression and epigenetics. Recent work developing an integrated view of genomics has found widespread interaction between the genome and these various levels (ENCODE Project Consortium et al. 2012). Better understanding of epigenetic inheritance, both theoretically and empirically, should help our understanding of this complex web of interactions.

10.2 Modeling Epigenetic Inheritance in a Population

A small but growing body of literature constructs and analyzes models of epigenetic inheritance. Several of these contributions focus on epigenetic impacts on phenotypic inheritance (Slatkin 2009; Danchin and Wagner 2010; Tal et al. 2010; Furrow et al. 2011; Day and Bonduriansky 2011). All of these models are built around the same fundamental concepts: an epigenetic state can be thought of as similar to a genetic or phenotypic trait, and a particular site in the genome can be in one of a set of possible epigenetic states. For example, a cytosine at a particular CpG site in the genome can be methylated or unmethylated. Or the state in question could be the

presence or absence of a small interfering RNA or other micro RNAs. The models therefore consider one or more epigenetic sites and are completed by specifying the mode of inheritance, described by parameters quantifying the probabilities that an epigenetic site switches from one state to another between generations. These switching probabilities may depend on the current state, the genotype, and the environmental condition, although simplifications are usually needed for clear analysis. When the focus of a model is epigenetic inheritance, the main emphasis of analysis is on the contrast between comparable models of epigenetic and genetic inheritance.

For example, Slatkin (2009) considered an arbitrary number of epigenetic loci, where the epigenetic states were transmitted like genetic states, except that the rates of epigenetic switching were higher than those typical of genetic mutation. The model assumed that the particular epigenetic state at each of the loci affected the overall disease risk of an individual. Slatkin concluded that higher rates of epigenetic switching between generations greatly reduced the contribution of the variation at these epigenetic loci to heritability. Similarly, the analysis of Tal et al. (2010) focused largely on the consequences of nontrivial rates of epigenetic switching (which they called resetting), in a manner analogous to genetic mutation.

The switching probabilities in an epigenetic model can be viewed as describing the inheritance of epigenetic modifications or as rates analogous to those of genetic mutations. In models of genetic inheritance, the environment has no effect on mutation rates unless mutagenic agents are explicitly included. But it is well understood, in a variety of organisms, that environmental conditions such as stress, diet, and temperature can influence the epigenetic states of individuals (Carone et al. 2010; Heijmans et al. 2008; Kucharski et al. 2008; McGowan et al. 2009; Ng et al. 2010; Sandovici et al. 2011; Tobi et al. 2009; Verhoeven et al. 2010; Waterland and Jirtle 2003; Weaver et al. 2004). Furthermore, the rates of “epimutation”—the probability of an epigenetic state switching during transmission between generations—can be several orders of magnitude higher than those expected for a genetic locus.

The correlation between the epigenetic state of a parent and its offspring will depend on the environments experienced by the individuals. Figure 10.2 shows how adults and their adult offspring may have high correlation in epigenetic state, even if there is significant resetting at some point during gametogenesis. If adults in a given environment produce offspring that are likely to remain in the same environment experienced by the adults, then they may have the same epigenetic marks reinduced despite low fidelity transmission of the epigenetic states to the zygotes produced by these adults.

General application of these population-based epigenetic inheritance models is currently limited by a shortage of transgenerational epigenetic data. Johannes and Colomé-Tatché (2011), however, used inheritance data from an experiment with the flowering plant *Arabidopsis thaliana*, where epigenetic variation had been induced while genetic variation was reduced (Johannes et al. 2009). Crosses were made between inbred lines with minimal genetic divergence but differences in CpG methylation rates. Their data analysis rested on a quantitative genetic model incorporating epigenetic variation, switching, and transgenerational transmission.

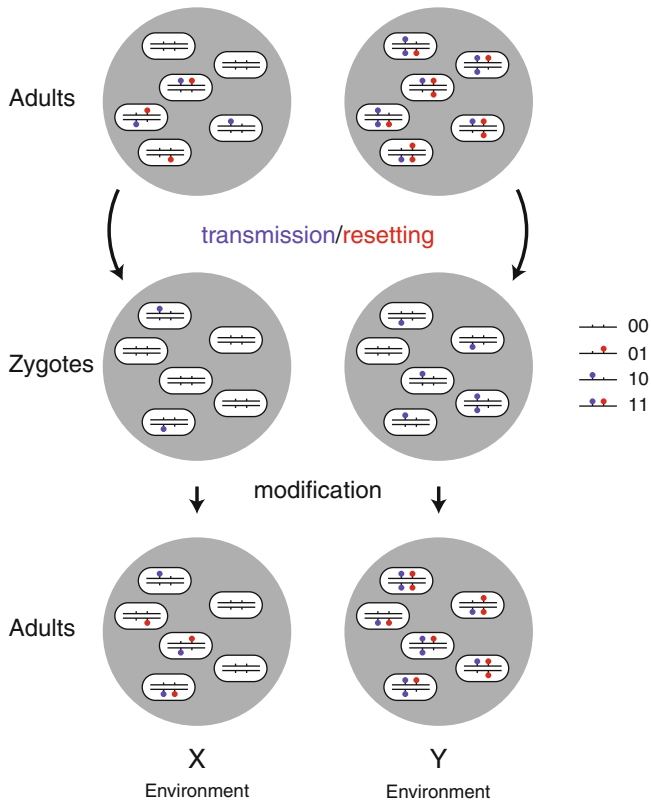


Fig. 10.2 The combined effects of low fidelity epigenetic transmission and shared environmental effects. Five individuals are shown in two sections of a stratified population living in environments that cause different epigenetic modifications. Reset at the site with *red* modifications is complete whereas two-third of the *blue* sites are transmitted. Although the epigenetic marks may be lost during production of the fetus, if an offspring is likely to experience a similar environment to its parents, then the shared environmental influence on rates of epigenetic switching can produce familial correlations in epigenetic state

They found high heritability in various quantitative traits despite the lack of genetic variance, and their analysis yielded estimates of the number of epigenetic sites contributing to variation.

10.3 Epigenetics and Phenotypic Heritability: Theoretical Considerations

To connect transgenerational epigenetic inheritance to phenotypic correlations between relatives, we must specify the influence of the epigenetic states on the phenotype in question. For a continuous quantitative trait, an epigenetic state may

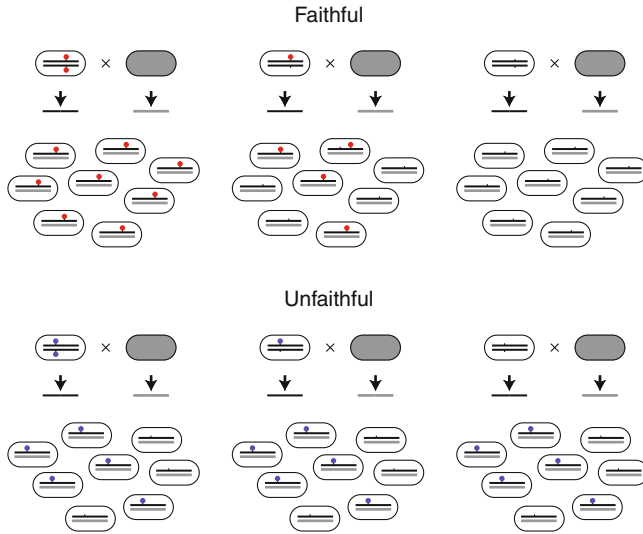


Fig. 10.3 Faithful and unfaithful epigenetic transmission. *Above:* Faithful transmission of an epigenetic modification (*red mark*) is illustrated in three crosses each with eight offspring. The epigenotype of a focal parent is shown, while the other parent is anonymous (*gray*). *Below:* The epigenetic modification considered (*blue mark*) may be lost between generations, or may arise de novo in the offspring. In this case there is no parent–offspring resemblance in the population

have some quantifiable influence, similar to the effect of allelic variation at a genetic locus in quantitative genetics (for example, as in Tal et al. 2010). For a discrete phenotype the epigenetic state can influence a continuous trait such as the risk of a particular disease (as in Slatkin 2009; Furrow et al. 2011) or some underlying liability trait that causes a different phenotype if the liability passes some threshold. The amount of epigenetic contribution to heritability of the phenotype in question depends on the relative importance of epigenetic versus genetic or environmental effects, and the magnitude of the covariance between the epigenetic states of parents and their offspring. As such, we focus on the factors influencing the familial covariance in epigenetic state.

Whatever the dynamic model we use to describe epigenetic transmission between generations, the epigenetic effect on heritability boils down to this covariance. The covariance is not a parameter of the dynamic model but a derived statistical property of the system, and a high covariance can occur in many different ways. For example, an epigenetic state that is transmitted very faithfully between generations may result in a high covariance, particularly when there is substantial epigenetic variation in the population (Fig. 10.3). Or a high covariance may result when different environmental states strongly induce particular epigenetic states, and the environments of parent and offspring are highly correlated (Fig. 10.2). In the first case, our epigenetic site behaves like a genetic locus with different alleles. But in the second, the epigenetic state is a manifestation of a heritable environment.

The epigenetic state need not be transmitted meiotically with high fidelity, as long as the induction of epigenetic changes is similar between parent and offspring.

Suppose that an individual's phenotype, P , can be expressed in terms of an epigenetic contribution Ω (borrowing notation from Johannes et al. 2009), and a non-epigenetic contribution N . The non-epigenetic contribution may include both genetic and environmental effects. If we use the subscript A for a focal adult and O for one of its offspring, then we can express an adult phenotype and that of one of its offspring as

$$P_A = \Omega_A + N_A \quad \text{and} \quad P_O = \Omega_O + N_O,$$

and express the covariance between the two individuals as

$$\begin{aligned} \text{Cov}(P_A, P_O) &= \text{Cov}(\Omega_A + N_A, \Omega_O + N_O) \\ &= \text{Cov}(\Omega_A, \Omega_O) + \text{Cov}(N_A, \Omega_O) + \text{Cov}(\Omega_A, N_O) + \text{Cov}(N_A, N_O). \end{aligned} \quad (10.1)$$

In this framework, phenotypic similarities between relatives could stem purely from shared genetics or environment, captured in the term $\text{Cov}(N_A, N_O)$. There may also be epigenetic contributions purely through direct transmission, $\text{Cov}(\Omega_A, \Omega_O)$. But epigenetic contributions to phenotypic covariance may also stem from associations between environments or genetics and epigenetic states (the terms $\text{Cov}(N_A, \Omega_O) + \text{Cov}(\Omega_A, N_O)$). Note that Eq. (10.1) does not explicitly incorporate any interaction terms between epigenetics, genetics, and the environment. It simply allows for covariances between the different phenotypic contributions across generations.

Initial theoretical work focused on epigenetic inheritance without incorporating variation in the environment. In this case, epigenetic variation is nothing more than genetic variation with high mutation rates. Accordingly, such studies compared $\text{Cov}(\Omega_A, \Omega_O)$ and $\text{Cov}(N_A, N_O)$ and found that the fidelity of epigenetic transmission between generations crucially determined the epigenetic contribution to heritability. Models of Slatkin (2009) and Tal et al. (2010) considered one or more epigenetic sites, with a fixed probability that the state will be reset during transmission from parent to offspring. By varying the parameter corresponding to rate of reset, they found that the epigenetic contribution to heritability decreased quickly as resetting increased. Only very low epigenetic reset rates would lead to a significant epigenetic contribution to the heritability of a phenotype (Slatkin 2009; Tal et al. 2010). Some empirical research has suggested that transgenerational inheritance of epigenetic states is unlikely in mammals, due to the widespread loss of methylation and other markers in the genome during gametogenesis, but exceptions to this effect have recently been noted (reviewed in Daxinger and Whitelaw 2012).

If the environment can influence rates of epigenetic switching, shared environments can mimic the effects of faithful meiotic transmission of epigenetic states. Furrow et al. (2011) modeled such a population and found several different

scenarios in which epigenetic inheritance could significantly contribute to phenotypic heritability: (1) when rates of epigenetic reset were low between parent and offspring, (2) when environmental correlations were high between family members, and (3) when both reset rates and environmental correlations were intermediate between parent and offspring. This model tracked individuals' epigenetic states, environmental states, and the states' transmission to offspring. The process of reproduction allowed for meiotic reset of states, in addition to induction of epigenetic state changes due to environmental influence, as demonstrated in Fig. 10.2. If rates of epigenetic reset were low regardless of the environment, then the Furrow et al. (2011) model was effectively identical to that of Slatkin (2009) or Tal et al. (2010). But the possibility of environmental correlations between parent and offspring allowed for new routes to high familial aggregations of epigenetic state. With high environmental correlations, and environmental states that caused different rates of epigenetic induction, correlations between the epigenetic states of parent and offspring could be high. These epigenetic correlations were especially high when individuals mated assortatively with respect to the environmental state they experienced. Furthermore, high contributions to heritability did not even require low epigenetic reset or high environmental correlations. Partial environmental correlations in conjunction with moderately faithful meiotic epigenetic transmission also yielded high epigenetic correlations between parent and offspring. Note that this model actually focused entirely on influences contained within the first summand in Eq. (10.1): there was no direct environmental influence on phenotype and hence no N term. In models that allow this environmental influence as well, we can expect environmental correlations to strengthen the epigenetic influence on phenotypic heritability even further. It would be possible to have significant contributions to phenotypic covariance from all four of the summands in Eq. (10.1). In this case, epigenetic contributions to phenotypic heritability are better seen as the combined effects of partial meiotic transmission and shared environments between generations. This insight applies equally well to classical analyses of cultural inheritance in circumstances where the environment cannot be controlled—which is the rule whenever human populations are investigated (Cavalli-Sforza and Feldman 1973, 1981; Feldman and Cavalli-Sforza 1979; Feldman et al. 1995, 2000; Otto et al. 1995).

Although it is never trivial to evaluate the environmental influence in such models, organisms amenable to experimentation can shed light on the possibilities. A study by Verhoeven et al. (2010) on apomictic dandelions tracked the effects of environmental stresses. Epigenetic changes were observed in the exposed individuals, and the homologous epigenetic sites were observed in genetically identical offspring. In response to stress, in particular chemical induction of herbivore or pathogen defences, plants showed widespread epigenetic changes in the genome, and many of these changes persisted in offspring raised from seeds in an environment lacking the stress. Such observations may allow estimates of environment-dependent epimutation rates, but because the phenomena may reflect a combination of many of biological processes, the conclusions drawn from one species may not apply to distantly related species.

10.4 Epigenetics and Phenotypic Heritability: Examples in Human Populations

Very little work has simultaneously tracked environments, epigenetics, and phenotypes in human populations. However, human populations have occasionally experienced strong, unexpected stresses that offer some insight into the potential interactions among environment, epigenetics, and phenotype. The stress of a famine, for instance, can strongly influence the phenotypes of individuals gestating or born during such a period. During World War II, the combination of a freezing cold winter and the German blockade of parts of the Netherlands led to a famine in the winter of 1944–1945, a period referred to as the Dutch Hungerwinter. Individuals prenatally exposed to this famine showed a significant increase in their risk of acquiring schizophrenia (Susser and Lin 1992; Susser et al. 1996), and even decades later, these individuals show epigenetic differences from their same sex siblings (Heijmans et al. 2008), although the effect appears to depend on both sex and gestational timing (Tobi et al. 2009). A spike in schizophrenia risk occurred among individuals conceived during the Chinese famine of 1959–1961 as well (St Clair et al. 2005). Obesity and other health risks are also associated with the Dutch Hungerwinter (Roseboom et al. 2006), and similar effects are associated with the famine in Biafra during the Nigerian civil war of 1967–1970 (Hult et al. 2010).

Studies in other mammals have demonstrated gestational dietary effects on the epigenetic profile of offspring. A study of sheep found that maternal undernutrition was associated with epigenetic changes in both CpG methylation and histone modifications in fetal hypothalamic pathways (Begum et al. 2012). Methylation rates in mice are affected by both maternal (Waterland and Jirtle 2003) and paternal diet (Carone et al. 2010; Ng et al. 2010), and paternal effects in mammals have been found to influence even the development of grand-offspring (Curley et al. 2011). It seems likely that at least some of the observed human phenotypic effects of the famines mentioned above may be mediated by environmentally induced epigenetic changes, as appears to be the case with other examples of stress in humans (Borghol et al. 2012; McGowan et al. 2009; Tyrka et al. 2012; Uddin et al. 2010; Waterland et al. 2010).

10.5 Epigenetics and Phenotypic Evolution

Another suite of models has focused on the evolutionary consequences of epigenetic inheritance due to adaptive variation in epigenetically controlled phenotypes that may vary between generations (Lachmann and Jablonka 1996; Thattai and van Oudenaarden 2004; Kussell and Leibler 2005; Salathe et al. 2009; Gaál et al. 2010; Feinberg and Irizarry 2010; Day and Bonduriansky 2011; Liberman et al. 2011; Carja and Feldman 2012; Geoghegan and Spencer 2012). In essence, these models extend investigations into phenotypic heritability by considering fitness as the phenotype in question. Some models focus on epigenetic sources of

fitness variation, while others consider the possibility of interaction between effects of epigenetic and genetic variation.

Day and Bonduriansky (2011) showed in a simple model of genetic and epigenetic interaction that rates of epigenetic reset between generations determine whether a population will evolve toward an equilibrium with epigenetic variation where the effects of selection balance with the rates of epigenetic reset, or toward an equilibrium in which there is no epigenetic variation. Geoghegan and Spencer (2012) found that there may be more than one equilibrium state stable in the population, i.e., different populations may show different levels of epigenetic variation without any differences in the environment or epigenetic transmission. In model populations where the rate of epigenetic switching can evolve, the evolutionarily stable epimutation rate will be related to the rate at which the environment changes. In all of these models the rates of epigenetic reset between generations (or the resulting correlation between epigenetic state of parent and offspring) determine the evolutionary dynamics of the epigenetically induced phenotypic variation, in particular the levels of variation at equilibrium with respect to selection and epigenetic reset between generations.

The level of phenotypic variation originating from epigenetic variation is maintained by the combination of environmental effects, selection, and rates of epigenetic reset between generations. However, the contribution of epigenetic variation to phenotypic inheritance is more narrowly determined by the direct transmission of epigenetic states to offspring and by the covariance between non-epigenetic influences and epigenetic influences on the phenotypes of parents and their offspring. The levels of epigenetic variation in the population may therefore be a weak indicator of the basis for phenotypic evolution, in particular for adaptation in response to selection. However, a study of a flowering plant found associations between browsing damage on the individual and its epigenetic status (Herrera and Bazaga 2011), and, within yeast populations, niche breadth and epigenetic variation were found to be correlated (Herrera et al. 2012). These studies suggest a possible role for epigenetic variation in short-term adaptation.

Although these results are promising, it is unclear whether a phenotype favored in a new environment will continue to increase in frequency for more than a few generations after the population arrived in that environment. An equilibrium maintained by a balance between selection and epigenetic dynamics may be far from the phenotypic optimum reached by analogous genetic systems, causing an “epimutational load” on the population. The evolutionary dynamics of environmentally induced epigenetic modifications may therefore be consistent with their ability to cause disease.

10.6 Measuring Epimutation Rates

One of the major challenges in empirical studies of epigenetics in the next decade will be the measurement of epigenetic switching rates and their variation through an organism’s lifespan and across different environments and tissues, for all types of

epigenetic marks (Fraga et al. 2005). Models of both evolution and phenotypic inheritance show that the fidelity of epigenetic transmission between generations is a critical parameter in inferences concerning either familial correlations or dynamics of adaptation. However, estimates of these rates are likely to depend on the age of the organisms. Inheritance of an epigenetic state from an adult to its juvenile offspring necessarily focuses on direct meiotic transmission and effects of shared environment early in life, but fails to illuminate the effects of shared environment that may be manifest on phenotypes measured later in life. This potentially leads to an underestimate of the importance of epigenetics in the inheritance of adult phenotypes, and to poor separation of the effects of direct transmission and common environment in early life.

10.7 Conclusion

Humans vary phenotypically due to many factors, some of which may be heritable, and some not. Epigenetic contributions to heritable phenotypic variation may stem from both genetic and environmental variation, in addition to random changes in the epigenetic state at some genomic positions. Epigenetic states may be directly transmitted through meiosis, thus mimicking genetic transmission with a high mutation rate. However, similar environments of parents and offspring may also produce phenotypic parent–offspring correlations in epigenetic states. In the study of transgenerational epigenetic inheritance care must therefore be taken to consider both the effects of direct transmission and those of the purely statistical associations between environment and epigenotype that emerges because of environmental influences on epigenotype. Both routes to transgenerational epigenetic inheritance have the potential to influence adaptation and phenotypic heritability, and both of these phenomena may explain some of the heritability not accounted for in genome-wide association studies of complex phenotypes and diseases (Eichler et al. 2010; Goldstein 2009; Maher 2008; Petronis 2010).

Epigenetic variation also offers a source of heritable phenotypic variation upon which natural selection can act. Epigenetic variation in a population should evolve toward a balance between selection and epigenetic modifications. But high rates of epigenetic modification, in an analogy to high mutation rates, will resist the adaptive push of natural selection. For epigenetic modifications to allow adaptation in static environments, we expect that the rates of epigenetic change at a site will be very low, akin to typical rates of genetic mutation.

As a body of theory is built, studies must revisit classic population genetics results, while relaxing assumptions about the magnitude of mutation rates and the influence of the environment on inheritance processes. Models of cultural evolution may also offer insights into the possible role of epigenetic inheritance in phenotypic evolution. At the same time, empirical research will clarify reasonable parameter ranges and assumptions. Only through the iterative process of improving both models and empirical methods we can begin to understand the full role of epigenetic variation in phenotypic heritability and evolution.

References

- Begum G et al (2012) Epigenetic changes in fetal hypothalamic energy regulating pathways are associated with maternal undernutrition and twinning. *FASEB J*. doi:10.1096/fj.11-198762
- Borghol N et al (2012) Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* 41(1):62–74
- Carja O, Feldman MW (2012) An equilibrium for phenotypic variance in fluctuating environments owing to epigenetics. *J R Soc Interface*. doi:10.1098/rsif.2011.0390
- Carone BR et al (2010) Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*. doi:10.1016/j.cell.2010.12.008
- Cavalli-Sforza LL, Feldman MW (1973) Cultural versus biological inheritance: phenotypic transmission from parents to children (a theory of the effect of parental phenotypes on children's phenotypes). *Am J Hum Genet* 25(6):618–637
- Cavalli-Sforza LL, Feldman MW (1981) Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton
- Curley JP et al (2011) Epigenetics and the origins of paternal effects. *Horm Behav*. doi:10.1016/j.yhbeh.2010.06.018
- Danchin E, Wagner RH (2010) Inclusive heritability: combining genetic and non-genetic information to study animal behavior and culture. *Oikos* 119(2):210–218
- Darwin C (1859) On the origin of species by means of natural selection. John Murray, London
- Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet*. doi:10.1038/nrg3188
- Day T, Bonduriansky R (2011) A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. *Am Nat*. doi:10.1086/660911
- Eichler EE et al (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. doi:10.1038/nrg2809
- ENCODE Project Consortium et al (2012) An integrated encyclopedia of dna elements in the human genome. *Nature*. doi:10.1038/nature11247
- Feinberg AP, Irizarry RA (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.0906183107
- Feldman MW, Cavalli-Sforza LL (1979) Aspects of variance and covariance analysis with cultural inheritance. *Theor Popul Biol* 15(3):276–307
- Feldman MW et al (1995) Statistics of discrete-valued traits under vertical transmission. Working paper No. 65 of the Morrison Institute for Population Studies—Stanford University. Accessible on web at <http://www.stanford.edu/group/morrinst/pdf/65.pdf>
- Feldman MW et al (2000) Genes, culture, and inequality. In: Arrow K, Bowles S, Durlauf S (eds) *Meritocracy and economic inequality*. Princeton University Press, Princeton
- Fraga MF et al (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.0500398102
- Furrow RE et al (2011) Environment-sensitive epigenetics and the heritability of complex diseases. *Genetics*. doi:10.1534/genetics.111.131912
- Gaál B et al (2010) Exact results for the evolution of stochastic switching in variable asymmetric environments. *Genetics*. doi:10.1534/genetics.109.113431
- Geoghegan JL, Spencer HG (2012) Population-epigenetic models of selection. *Theor Popul Biol*. doi:10.1016/j.tpb.2011.08.001
- Goldstein DB (2009) Common genetic variation and human traits. *N Engl J Med*. doi:10.1056/NEJMp0806284
- Heijmans BT et al (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.0806560105
- Herrera CM, Bazaga P (2011) Untangling individual variation in natural populations: ecological, genetic and epigenetic correlates of long-term inequality in herbivory. *Mol Ecol*. doi:10.1111/j.1365-294X.2011.05026.x

- Herrera CM et al (2012) Jack of all nectars, master of most: Dna methylation and the epigenetic basis of niche width in a flower-living yeast. *Mol Ecol*. doi:10.1111/j.1365-294X.2011.05402.x
- Hult M et al (2010) Hypertension, diabetes and overweight: looming legacies of the bialfran famine. *PLoS ONE*. doi:10.1371/journal.pone.0013582
- Johannes F, Colomé-Tatché M (2011) Quantitative epigenetics through epigenomic perturbation of isogenic lines. *Genetics*. doi:10.1534/genetics.111.127118
- Johannes F et al (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet*. doi:10.1371/journal.pgen.1000530
- Johannsen W (1903) Om arvelighed i samfund og i rene linier. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger* 3:247–270
- Kucharski R et al (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science*. doi:10.1126/science.1153069
- Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. doi:10.1126/science.1114383
- Lachmann M, Jablonka E (1996) The inheritance of phenotypes: an adaptation to fluctuating environments. *J Theor Biol*. doi:10.1006/jtbi.1996.0109
- Lieberman U et al (2011) On the evolution of mutation in changing environments: recombination and phenotypic switching. *Genetics*. doi:10.1534/genetics.110.123620
- Maher B (2008) Personal genomes: the case of the missing heritability. *Nature*. doi:10.1038/456018a
- McGowan PO et al (2009) Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci*. doi:10.1038/nn.2270
- Ng SF et al (2010) Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* 467(7318):963–967
- Otto SP et al (1995) Genetics and cultural inheritance of continuous traits. Working paper No. 64 of the Morrison Institute for Population Studies—Stanford University. Accessible on web at <http://www.stanford.edu/group/morrinst/pdf/64.pdf>
- Petronis A (2010) Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*. doi:10.1038/nature09230
- Roseboom T et al (2006) The Dutch famine and its long-term consequences for adult health. *Early Hum Dev*. doi:10.1016/j.earlhumdev.2006.07.001
- Salathe M et al (2009) Evolution of stochastic switching rates in asymmetric fitness landscapes. *Genetics*. doi:10.1534/genetics.109.103333
- Sandovici I et al (2011) Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the *hnf4a* gene in rat pancreatic islets. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1019007108
- Skinner MK (2011) Role of epigenetics in developmental biology and transgenerational inheritance. *Birth Defects Res C Embryo Today*. doi:10.1002/bdrc.20199
- Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics*. doi:10.1534/genetics.109.102798
- St-Clair D et al (2005) Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959–1961. *JAMA*. doi:10.1001/jama.294.5.557
- Susser ES, Lin SP (1992) Schizophrenia after prenatal exposure to the Dutch hunger winter of 1944–1945. *Arch Gen Psychiatry* 49(12):983–988
- Susser ES et al (1996) Schizophrenia after prenatal famine—further evidence. *Arch Gen Psychiatry* 53(1):25–31
- Tal O et al (2010) Epigenetic contribution to covariance between relatives. *Genetics* 184(4):1037–1050
- Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. *Genetics* 167(1):523–530
- Tobi EW et al (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet* 18(21):4046–4053
- Tyrka AR et al (2012) Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. *PLoS ONE* 7(1):e30148

- Uddin M et al (2010) Epigenetic and immune function profiles associated with posttraumatic stress disorder. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.0910794107
- Verhoeven KJF et al (2010) Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol*. doi:10.1111/j.1469-8137.2009.03121.x
- Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23(15):5293–5300
- Waterland RA et al (2010) Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* 6(12):e1001252
- Weaver ICG et al (2004) Epigenetic programming by maternal behavior. *Nat Neurosci* 7(8):847–854

Chapter 11

Statistical Approaches for Detecting Transgenerational Genetic Effects in Humans

Janet S. Sinsheimer and Michelle M. Creek

Abstract Transgenerational genetic effects occur when the genes of one generation influence the phenotype of subsequent generations without Mendelian transmission of alleles, possibly through inherited epigenetic effects. The evidence for transgenerational genetic effects in humans comes predominantly from genetic epidemiology studies, which thus presents a number of statistical challenges to their analysis and interpretation. In this chapter, we outline some of the genetic epidemiologic study designs and statistical analysis approaches that have been used to detect these effects and discuss their strengths and weaknesses.

11.1 Introduction

Genetic epidemiology concentrates on disease risks due to a subject's own genes and environment. Although we gain much etiological insight from these studies, many genetic determinants of disease remain undiscovered. One possibility is that transgenerational genetic effects play a role in their etiology. Transgenerational genetic effects occur when the genes of one generation influence the phenotype of subsequent generations without Mendelian transmission of alleles (Fig. 11.1), possibly through inherited epigenetic effects (Gluckman et al. 2007; Nadeau 2009). Most commonly, these transgenerational genetic effects are parental genes having an effect on their offspring's phenotype, but more distant ancestors can have effects.

J.S. Sinsheimer (✉)

Department of Human Genetics, David Geffen School of Medicine at UCLA,
Los Angeles, CA 90095, USA
e-mail: janet@mednet.ucla.edu

M.M. Creek

Department of Biostatistics, UCLA Fielding School of Public Health,
Los Angeles, CA 90095, USA
e-mail: mmcreek@gmail.com

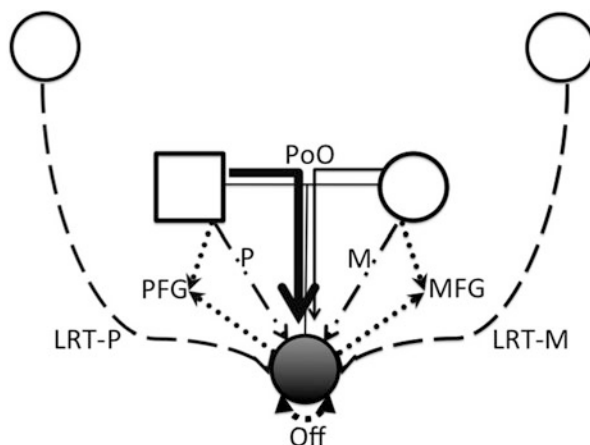


Fig. 11.1 Schematic depiction of potentially detectable effects. Disease risk in the offspring (denoted by the *dark circle*) can be due to maternal (M), paternal (P), offspring (Off), parent of origin (PoO), long-range transgenerational (LRT-M maternal side, LRT-P paternal side), maternal–fetal genotype incompatibility (MFG), or paternal–fetal genotype incompatibility (PFG) effects. These effects are not mutually exclusive. The effects could be genetic or environmental in origin. When genetic in origin, they have direct effects or are epigenetic or environmentally mediated

Less commonly considered, an offspring's genotype could also elicit a phenotype in his/her mother. In this chapter, we explore statistical approaches for detecting transgenerational genetic effects in humans with genetic epidemiological data. With a few exceptions, these studies provide only indirect evidence consistent with epigenetic phenomena and, in some of these cases, the underlying explanations for transgenerational genetic effects will not involve modifications to DNA or chromatin. However, we argue these studies provide an excellent starting point for hypothesis generation and for further investigations leading to more direct tests for epigenetic effects.

To clarify what we mean by transgenerational genetic effects, we first provide descriptions of some common ones before going on to describe appropriate study designs and analysis approaches to detect them. For notational convenience we drop the “genetic” and from now on refer to them collectively as transgenerational effects. Figure 11.1 illustrates some transgenerational effects discussed in this chapter including maternal effects (M), paternal effects (P), parent of origin effects (PoO), maternal–fetal genotype (MFG) incompatibility, paternal–fetal genotype (PFG) incompatibility, and long-range transgenerational effects from the maternal or paternal side (LRT-M or LRT-P).

Maternal genetic effects and paternal genetic effects. There are observations that particular maternal genotypes are strongly associated with offspring phenotypes, regardless of what alleles the offspring inherits (Kistner and Weinberg 2004; Wheeler and Cordell 2007; Weinberg et al. 1998). Although discussed less often, paternal genotypes are also associated with offspring phenotype. A possible mechanism for

these effects includes parental genotype influences on the offspring's environment. This environmental effect could have no effect on the offspring's DNA or chromatin or it could induce epigenetic modifications. Alternatively, parental genotype could directly affect the offspring's epigenome or the effects could be due to maternally derived mitochondrial DNA.

Parent of origin effects. PoO effects occur when the degree of association of an allele with an offspring's phenotype depends on the sex of the transmitting parent (in Fig. 11.1, the effects of the paternally transmitted allele prevail). Although silencing through methylation or histone modification, commonly referred to as imprinting, is one form of PoO effects, other mechanisms can also lead to these effects (Guilmatre and Sharp 2012). These mechanisms include mutational transmission bias and oocyte RNAs or proteins.

MFG incompatibility. The effects of maternal genes on the offspring's disease risk may vary depending on the offspring's genotype (Fig. 11.1). MFG incompatibilities are gene interactions that produce adverse effects on the developing fetus. These gene–gene interactions differ from typical ones because maternal genes interact with offspring genes. MFG incompatibilities are involved in complex diseases, even adult onset diseases where the effects may not be evident until long after the MFG incompatibility initiated event has occurred and subsided (Palmer et al. 2002, 2006; Sinsheimer et al. 2003). In principle, there could be paternal–offspring gene interactions (PFG incompatibility), although there is less biological support for these interactions than for MFG incompatibility.

Like PoO and maternal effects, the mechanisms by which MFG or PFG incompatibilities occur could be methylation or chromatin modification but other mechanisms are possible. The prototypical MFG incompatibility is *RHD* incompatibility, which can lead to erythroblastosis, liver damage, hypoxia, or death from hemolytic disease of the newborn (HDN) (Guyton 1981). The biological mechanism underlying *RHD*-induced HDN is well known (Stratchen and Reed 2003) and we provide a simplified description. Alleles at the *RHD* locus are classified into two types, *D* and *d*. The *D* allele codes for an antigen on the erythrocyte surface and the *d* allele is a null allele. *RHD*-induced HDN occurs when a mother with a null allele homozygous genotype (*d/d*) mounts an IgG alloimmune response to her *d/D* offspring's erythrocytes, damaging their ability to carry oxygen and releasing bilirubin. Maternal–fetal *ABO* incompatibility leads to HDN by a similar mechanism (Guyton 1981). *RHD* and *ABO* incompatibilities are implicated as risk factors for complex diseases (Cannon et al. 2002; Dahlquist et al. 1999; Hollister et al. 1996; Insel et al. 2005; Juul-Dam et al. 2001; Kraft et al. 2004; Palmer et al. 2002; Stubbs et al. 1985). Although *RHD* incompatibility involves the same locus in mother and offspring, MFG incompatibilities can also occur between one locus in the mother and another locus in the offspring (Chen et al. 2009).

Long-range transgenerational (LRT) effects. It is difficult to distinguish epigenetic effects from shared environment unless the transgenerational effect persists over multiple generations but the environmental exposure does not. Environmental exposures, even if they are short lived, can affect three generations without

involving specific inherited epigenetic factors. If a pregnant woman is exposed to an environmental stimulus, she, her fetus, her gametes, and her fetus' gametes can be directly affected without involving epigenetic modifications. Likewise environment can affect two generations when a man is exposed because his gametes can be affected. LRT genetic effects caused by inherited epigenetic effects are well documented in model organisms (Nadeau 2009), and there is evidence of their role in common diseases in humans (e.g., Benyshek et al. 2001; Klip et al. 2002).

11.2 Study Designs

In this section, we discuss epidemiological study designs used to detect transgenerational effects. Most statistical approaches to study these effects have been designed for bivariate, qualitative traits. Therefore, when we discuss specific study designs and analyses, we concentrate on these bivariate traits and refer to cases and controls. When methods for continuous traits are commonplace we discuss them in the analysis section along with the appropriate modifications to the study design.

Case-mother, control-mother (CMCM). The CMCM design allows detection of offspring genetic effects, maternal genetic effects, and their interactions by comparing the genotype distributions of affected individuals and their mothers to the genotype distributions of unaffected individuals and their mothers (Ainsworth et al. 2011). This design is an extension of the popular case-control design of genome-wide association and is subject to the same limitations, such as confounding from population substructure.

Case-parent trios (CPTs). CPTs were first popularized in genetics to avoid the problems of population substructure that originally plagued case-control genetic analysis (e.g., Laird and Lange 2010). Although this advantage is largely eliminated by methods that control for ancestry in case-control studies (e.g., Edwards and Gao 2012), CPTs are popular for detecting transgenerational effects associated with disease (Cordell 2004; Cordell et al. 2004; Laird and Lange 2010). Using CPTs expands the genetic models that can be considered over using CMCMs. For example one can test for PoO and paternal effects. For a bi-allelic locus there are 15 possible maternal-paternal-offspring genotype combinations. Case-mother and case-father duos can be included along with the CPTs by treating the duos as trios with randomly missing data (Sinsheimer et al. 2003; Weinberg et al. 1998). These models can be modified to include control-mother duos or parents of unaffected offspring (Vermeulen et al. 2009) but then population stratification comes back into play.

Nuclear families. The CPT design can be extended to include unaffected and affected siblings of the case (Kraft et al. 2004). These extensions provide additional power and increase the genetic models that can be considered but may require additional modeling assumptions or else be biased. As we discuss in the statistical analysis section, the study design dictates the questions that can be posed as well as the assumptions imposed.

General pedigree data. Intuitively, the ideal study design for detecting transgenerational effects allows simultaneous analysis of unrelated individuals, small pedigrees, and large pedigrees. The inclusion of large, multigenerational pedigrees provides a way to study a variety of complex patterns and detect LRT effects. Being able to analyze all family members is highly efficient. Breaking up large pedigrees into subsets can introduce bias (Childs et al. 2010, 2011). Pedigrees provide a way to model phenotype data in the absence of genotype data (see the statistical analysis section). One disadvantage is that, depending on the research question, using multigenerational families requires more restrictive modeling assumptions to be computationally feasible.

11.3 Statistical Analysis Approaches for Detecting Transgenerational Effects

We briefly outline some statistical analysis approaches. Because the approaches depend on the available data, we group them by data type: (1) phenotype data only, (2) phenotype and genotype data, and (3) phenotype, genotype, and epigenetic data.

11.3.1 Approaches Using Only Phenotype Data

Prior to the wide spread availability of genotype data, evidence supporting the existence of transgenerational effects in humans came from the inference of phenotypic inheritance patterns inconsistent with Mendelian inheritance. These approaches generally require large pedigrees to be effective but, with marked environmental exposures, following matrilineal or patrilineal lines provides evidence of transgenerational effects (e.g., Gluckman et al. 2007).

Indirect evidence for transgenerational effects can, in principle, be obtained from analyzing pedigrees with complex segregation analyses (e.g., Khoury et al. 1993). These analyses use correlations among family members' phenotypes to infer the existence of major genes acting in a Mendelian manner, polygenes, shared environment, and independent environment (residual effects). Generational differences and birth order effects can be inferred. The number of effects inferred is dependent on the variety of relationships and so, in general, large pedigrees are needed to adequately explore transgenerational effects. The biggest difficulties with this approach are the equivalence or near equivalence of sets of models and the inability to prove any model to be true.

Variance component analysis (e.g., Lange 2002) and its related approach, path analysis (e.g., Thomas 2004) have been used to separate genetic sources of phenotypic variation from other sources. In the absence of genetic marker data, what is not the effect of a gene or genes (possibly many) acting in a Mendelian fashion is typically assumed to be environmentally induced. These approaches postulate trait

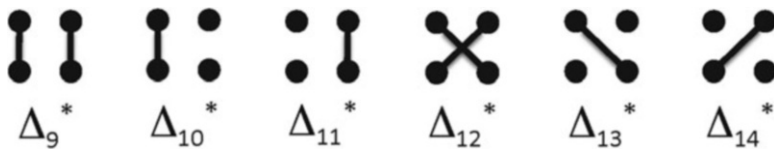


Fig. 11.2 Detailed identity states used in the variance component analyses. Six detailed identity states describe the IBD sharing between two non-inbred individuals i and j . Four circles in a block represent the genes from i and j at a single locus. Individual i 's genes are on the *top* and j 's genes are on the *bottom*. Maternally derived genes are on the *left*, paternally derived genes on the *right*. Lines between genes represent genes that are identical by descent. The probability of observing state k is denoted by Δ_k^*

variation is partitioned. Classically, genetic effects are modeled as many genes acting approximately equally and independently (polygenes). The additive genetic variance captures the effect of alleles at these genes as if they were acting independently, deviation from allelic independence leads to dominance genetic variation. The genetic correlation between two relatives' trait values depends on the expected distribution of genes shared identically by descent (IBD) among them (Fig. 11.2). Environmental variation is postulated to come in two forms: shared and independent among the pedigree members. Shared environmental variation captures the additional correlations among family members that remain unexplained by IBD sharing. Adopted relatives and other unrelated individuals living in the same household help to distinguish these correlations from genetic correlations. Independent environmental variation is any residual variation in the trait values after accounting for genetic and shared environmental variations. It is always included because measurement errors make the model fit imperfect.

Although it is possible to use variance component models to test for transgenerational effects with only phenotypes, this approach has not been pursued to any appreciable extent. One reason is that shared environment and transgenerational effects are often confounded, making inference of transgenerational effects difficult. Parent of origin effects provide an exception (Gorlova et al. 2007; Zhou et al. 2011). PoO effects lead to a difference in parent–offspring correlations depending on the parent's sex and thus are accommodated by partitioning the additive genetic variance into two separate effects. When there are no shared environmental effects but there are parent of origin effects, the variance covariance matrix for family phenotypes Y can be written as:

$$\begin{aligned} \text{Var}(Y) = & (\Delta_9^* + \Delta_{10}^*)\sigma_{\text{ma}}^2 + (\Delta_9^* + \Delta_{11}^*)\sigma_{\text{pa}}^2 \\ & + (2\Delta_{12}^* + \Delta_{13}^* + \Delta_{14}^*)\text{cov}_{\text{mpa}} + \Delta_9^*\sigma_{\text{d}}^2 + \Delta_{12}^*\text{cov}_{\text{d}} + I\sigma_{\text{e}}^2, \end{aligned} \quad (11.1)$$

where σ_{ma}^2 is the maternal additive genetic variance, σ_{pa}^2 is the paternal additive genetic variance, cov_{mpa} is the additive covariance of the maternal and paternal alleles, σ_{d}^2 is the dominance genetic variance, cov_{d} is the dominance covariance, I is the identity matrix, σ_{e}^2 is the independent environmental variance, and Δ_i is the probability of IBD state i (Fig. 11.2).

Testing for maternal genetic effects with variance component models presents a problem, especially when attempting to dissect prenatal and postnatal effects from maternally inherited effects. This dissection is important because if prenatal effects are not properly modeled, heritability estimates are biased and may lead to false inference of PoO effects (Zhou et al. 2011). In the absence of measured predictors (e.g., genotypes), it is impossible to estimate all three effects in traditional nuclear families because the mother providing the genetic material (genetic mother) is the same person carrying the child (gestational mother) and the same person raising the child (postnatal mother). Although animal experimentation provides opportunities to dissect these effects by embryo transplantation and cross fostering (e.g., Nadeau 2009), in humans the options are limited. Adoption studies have been used to separate out postnatal effects but prenatal and maternal inherited effects are still confounded. Comparing the offspring of sisters to the offspring of brothers separates maternal inheritance from prenatal and postnatal effects (Robson 1955). However, recent advances in assisted reproductive technologies (ART) provide ways to separate all these effects (Thapar et al. 2007; Zhou et al. 2011).

Because ART uses sperm donation, egg donation, or gestational surrogacy depending on approach, children's genetic parents can be different from their prenatal or postnatal parents. By comparing phenotypes within and between these families, it is possible to separate maternal genetic, prenatal, and postnatal effects (Zhou et al. 2011). The key is to modify equation (11.1) by adding in a prenatal household matrix H_{pre} , where offspring born to the same gestational mother are indicated and a postnatal household matrix H_{post} where members with common environmental exposures are indicated.

$$\begin{aligned} \text{Var}(Y) = & (\Delta_9^* + \Delta_{10}^*)\sigma_{\text{ma}}^2 + (\Delta_9^* + \Delta_{11}^*)\sigma_{\text{pa}}^2 + (2\Delta_{12}^* + \Delta_{13}^* + \Delta_{14}^*)\text{cov}_{\text{mpa}} \\ & + \Delta_9^*\sigma_{\text{d}}^2 + \Delta_{12}^*\text{cov}_{\text{d}} + I\sigma_{\text{e}}^2 + H_{\text{pre}}\sigma_{\text{pre}}^2 + H_{\text{post}}\sigma_{\text{post}}^2. \end{aligned} \quad (11.2)$$

Phenotypes from all the parents and offspring in ART families can be used to estimate these effects. Known risk factors are included as fixed covariates. Zhou (Zhou et al. 2011) demonstrates this approach with nuclear families, and note that variance component models can, in principle, use ART pedigrees of any complexity.

11.3.2 Approaches Using Phenotype and Genotype Data

When genotypes are available, the possibilities for detecting transgenerational effects improve. Genotypes provide a causal anchor and allow dissection of genetic effects from environmental effects. Tests of the association of parental genotypes with offspring phenotype and interactions between parent and offspring genotypes are possible. Direct evidence for the epigenetic mechanisms underlying these transgenerational effects is not obtained from these studies; however, support for persistent, shared environment can be reduced or eliminated.

Table 11.1 Joint maternal–offspring genotype relative risks

Genotype		Model			
Mother ^a	Offspring	Joint maternal— offspring ^b	Main effects and interactions	<i>RHD</i> incompatibility ^c	NIMA ^d
2/2	2/2	δ_{22} ^e	S_2R_2 ^f	1	R_2
2/2	1/2	δ_{21}	$S_2R_1 \gamma_{21}$	1	R_1
1/2	2/2	δ_{12}	S_1R_2	1	R_2
1/2	1/2	δ_{11}	S_1R_1	1	R_1
1/2	1/1	δ_{10}	S_1	1	γ_{10}
1/1	1/2	δ_{01}	$R_1 \gamma_{01}$	γ_{01}	R_1
1/1	1/1	δ_{00}	1	1	1

^aThe first two columns denote genotypes of the mother and her offspring
^bColumns 3 and 4 denote two different parameterizations of the most general model of maternal–offspring genotype effects that can be used with a bi-allelic locus
^cColumn 5 models *RHD* incompatibility
^dColumn 6 models NIMA and offspring main effects
^e δ_{ij} is the joint effect of *i* maternal 2 alleles and *j* offspring 2 alleles
^f S_i is the main effect of *i* maternal 2 alleles R_j is the main effect of *j* offspring 2 alleles and γ_{ij} is the additional interaction effect

Before embarking on a discussion of the specific analysis approaches, we note that most statistical approaches assume that the genotype is a SNP. Thus we will also focus on a single bi-allelic locus. Readers should be aware however there are a few methods that allow multi-allelic and multi-locus genotype data (e.g., Chen et al. 2009; Childs et al. 2011; Hsieh et al. 2006a; Sinsheimer et al. 2003).

CMCM data can be summarized in a two factor contingency table and analyzed as chi-square or using a Fisher exact test. Analysis of CMCM data can incorporate covariates affecting disease susceptibility by using logistic regression. The combinations of mother–offspring genotypes represent levels of one factor and case–control status represents levels of a second factor. Thus these data can be used to test models regarding maternal genotype main effects, offspring genotype main effects, and their interactions. Under the null hypothesis of no effect of this locus on disease susceptibility, the two factors are independent and the genotype frequencies for cases should be the same as the genotype frequencies for controls.

The number of levels for the first factor depends on whether the maternal and offspring SNPs are at the same locus or two distinct loci. When the maternal and offspring loci are distinct, there are nine possible maternal–offspring combinations. Chen et al. (2009) proposed a likelihood ratio test that allows the inclusion of mother’s and offspring’s genotypes at both these loci and showed that including both increases overall information and thus increases power. When considering the same bi-allelic locus for mother and offspring, the number of maternal–offspring genotype combinations is seven (Table 11.1). Assumptions regarding the mechanisms by which these genotypes lead to disease or how the maternal and offspring genotypes interact result in restrictions on the levels, further reducing the number of independent parameters. If paternal–offspring interactions are suspected,

the CMCM design can be changed to a case-father control-father (CFCF) design and analyzed in the same manner.

Inference from CMCM studies is sensitive to population stratification. Although this can be corrected by accounting for maternal population history, CMCM designs are limited in the hypotheses that can be tested. One alternative is to use case-parent trios. Several analysis approaches have been used to analyze trios depending on the research questions under consideration.

Parent of origin effects can be detected using the transmission disequilibrium test (TDT) (Spielman et al. 1993; Terwilliger and Ott 1992). The TDT is a form of conditional logistic regression that uses a retrospective design where the genotype of the offspring is the dependent variable (Sham and Curtis 1995; Sinsheimer et al. 2000; Thomas 2004). The TDT is a test of linkage and association between a genetic marker and a disease locus. When used with CPTs, the null hypothesis is no linkage or no association and a heterozygous parent is equally likely to pass on either of their alleles to their offspring. If there are linkage and association, one allele will appear to be transmitted more often to an affected offspring than the other. By comparing a model allowing for separate maternal and paternal transmissions to the standard TDT where the maternal and paternal transmissions are the same, the existence of PoO effects can be tested.

The TDT can be further modified to examine parent-offspring genotype interactions. For example non-inherited maternal antigen (NIMA) effects, a form of MFG incompatibility postulated to occur in rheumatoid arthritis (Harney et al. 2003; Hsieh et al. 2006b), can be tested by comparing the proportion of cases whose genotypes are incompatible with their mother's genotype to the proportion of cases whose genotypes are incompatible with their father's genotype (Harney et al. 2003). The assumption underlying this analysis is that NIMA is a plausible risk factor for a complex disease, but non-inherited paternal antigens (NIPA) are not. One major deficit of this design is that it is not possible to simultaneously check for offspring genotype effects, and maternal genotype effects are confounded with MFG incompatibility. The design also requires a substantial number of fathers be genotyped to have reasonable power.

The TDT gains no information from parents with homozygous genotypes, limiting power. Weinberg (Weinberg et al. 1998) proposed a log-linear model as an alternative and tested for offspring genetic main effects, maternal genetic main effects, and parent of origin effects. Sinsheimer (Sinsheimer et al. 2003) recognized the log-linear model could be extended to allow for maternal-offspring gene interaction at a single locus. Like the TDT, these log-linear models use cases and their parents in a retrospective design in which the genotypes are the dependent variables and no controls are necessary. Sinsheimer's MFG test maximizes the equivalent multinomial likelihood to the log-linear model in order to estimate parameters, and thus easily accommodates maternal-offspring and paternal-offspring dyads as incomplete trios.

The MFG (and equivalently the log-linear) test is very flexible, allowing many inherited disease risk scenarios to be modeled (Ainsworth et al. 2011; Hsieh et al. 2006a, b, 2007; Minassian et al. 2006). When using CPTs and a single

bi-allelic locus, there are 15 possible offspring–maternal–paternal genotype combinations. Under the null model of no genetic effects on the phenotype, Mendelian transmission holds and the number of independent parameters reduces to eight, one less than the number of maternal–paternal genotype combinations (mating types). If one assumes the sex of the parent is irrelevant in determining the probability of the mating types (the symmetric mating assumption), then the nine combinations reduce to six. If random mating with regards to the locus holds, then the mating types can be parameterized in terms of the three genotype frequencies and leads to two independent parameters to estimate. The number of independent parameters under the null further reduces to one if Hardy Weinberg Equilibrium is assumed.

Maternal and offspring genotype effects are estimated as genotype relative risks along with mating-type frequencies. Table 11.1 presents the same mother–offspring combinations and genotype relative risks for a bi-allelic locus as can be modeled with CMCM data. Columns 3 and 4 present two mathematically equivalent parameterizations for the most general model of maternal–offspring effects with a bi-allelic locus. Although column 3, the joint risk model, has seven parameters, the maximum number of maternal–offspring parameters that can be estimated is six because one of these joint risks is the referent with value one. Column 4 is parameterized in terms of maternal main effects, offspring main effects, and two MFG incompatibilities. We also present two examples of restrictions. The model in column 5 represents *RHD* incompatibility without offspring or maternal main effects. Column 6 represents NIMA effects along with offspring genotype effects (Hsieh et al. 2007). All of these models are available for testing in the MFG option of the Mendel Statistical Genetics Software Package (Lange et al. 2013).

The log-linear and equivalent multinomial approaches can also test for the existence of PoO effects in the possible presence of maternal and offspring effects (Ainsworth et al. 2011; Weinberg et al. 1998). These authors caution against over-parameterization and discuss the problem of multiple interpretations.

Although CPTs have much to offer, many families have multiple-affected offspring and including only one of these offspring is inefficient. In order to use any number of affected siblings per family, Kraft et al. (2004) used a conditional retrospective likelihood approach. This approach finds the likelihood of the genotypes conditional on the affection status of the siblings and can estimate the offspring, maternal (or paternal) genotype effects, and their interactions by including these effects in the penetrance function. Families where one or both parents have missing genotypes are included in the likelihood by summing over all possible genotypes for the missing parents. Unaffected siblings are treated as phenotype unknown. The genotypes of these unaffected or phenotype unknown offspring can be included in the likelihood to help infer the possible genotypes for missing parents without introducing any bias provided the disease is not too common (Hsieh et al. 2006a). If the locus under study is causal, is unlinked to other causal loci and there are no gene–environment interactions, then the penetrance functions of the offspring are independent conditional on their own genotype and that of their mothers. The maximum likelihood estimates of the relative risks and the

$$L(\mathbf{G} | \mathbf{D}) = \frac{\sum_{g_1} \dots \sum_{g_n} \prod_j \text{Prior}(g_j) \prod_i \text{Pr}(G_i | g_i) \prod_{[c, r, s]} \text{Pr}(D_c | g_c, g_r) \text{Trans}(g_c | g_r, g_s)}{\sum_{g_1} \dots \sum_{g_n} \prod_j \text{Prior}(g_j) \prod_{[c, r, s]} \text{Pr}(D_c | g_c, g_r) \text{Trans}(g_c | g_r, g_s)}$$

Fig. 11.3 The retrospective likelihood for a single, arbitrary pedigree. With multigenerational pedigrees, the number of mating types becomes impractically large so founders' genotypes are used under an assumption of random mating. Founder j 's genotype frequency = $\text{Prior}(g_j)$. These frequencies are estimated along with the other parameters in the likelihood. With CPTs or two generation nuclear families, $\text{Prior}(g_j)$'s are replaced with mating-type frequencies $\text{MT}(g_r, g_s)$. $\text{Pr}(G_i | g_i) = 1$ if the proposed genotype for any pedigree member i , g_i , is consistent with the observed genotype G_i , and 0 otherwise. When G_i is missing, $\text{Pr}(G_i | g_i) = 1$. $\text{Pr}(D_c | g_c, g_r)$ is the offspring c 's disease probability dependent on both their and their mother's genotype. For computational ease, $\text{Pr}(D_c | g_c, g_r)$, is calculated with $\text{Trans}(g_c | g_r, g_s)$, the transmission probability for offspring, mother, and father triples (c, r, s) . The denominator sums over all possible ordered (phased) genotypes for the n family members. The likelihoods of independent pedigrees multiply. When there are only CPTs, the denominator is constant and is not relevant to the inference. The likelihood of the study samples is then proportional to a 15-mer multinomial

mating-type frequencies are obtained by solving score equations of the sample log likelihood and the standard errors of the estimates are derived through the observed information matrix (Lange 2002). Null hypotheses are tested using likelihood ratio test statistics.

Besides allowing more data to be used, an advantage of using nuclear families is that prior exposure effects can be tested. In this case, the genotypes of unaffected or phenotype unknown siblings fulfill an additional role of defining prior exposure. Kraft et al. (2004) used nuclear families to test whether risk of schizophrenia increased for offspring who were *RHD* incompatible when their older sibling was also *RHD* incompatible and found support for this hypothesis.

This conditional retrospective likelihood approach can be extended for use with large pedigrees. Like the nuclear family test, the extended MFG incompatibility (EMFG) test examines both maternal and offspring genotypes as risk factors for disease. The EMFG test jointly models maternal genotype effects, offspring genotype effects, and maternal–offspring genotype interactions using a retrospective likelihood (see Fig. 11.3 for mathematical details). Childs et al. (2010, 2011) developed this approach to allow any pedigree to be used including those with multiple generations and multiple-affected individuals. To reduce the number of nuisance parameters, the EMFG test replaces mating types with founder genotypes and assumes random mating with respect genotypes among the founders. The EMFG test handles multi-allelic loci, including non-codominant loci and several tightly linked loci, and can also incorporate potential offspring-related confounders. The EMFG test likelihood uses the classic formulation of the pedigree likelihood (Ott 1974) and modifies it by (1) conditioning on the phenotypes and (2) using penetrance functions that depend on both the offspring and maternal genotypes (Fig. 11.3). Each pedigree has its own conditional likelihood and these conditional likelihoods multiply. Unaffected family members are treated as phenotype unknown. Although EMFG is an affected-only analysis, the genotypes of the

unaffected or phenotype unknown offspring are used when there are missing parental genotypes.

The conditional likelihood used by Kraft et al. (2004) and Childs et al. (2010, 2011) has additional attractive features. Under relatively mild assumptions regarding the conditional independence of environmental exposure and offspring genotypes given parental genotypes, the effects of environmental covariates can be incorporated into the models. Serotypes and other non-codominant markers can be used by treating the genotypes underling these phenotypes as missing data (Minassian et al. 2006). The conditional likelihood approach also has some disadvantages. When the locus under study is not the causal locus but is linked to the causal locus, the variance in the parameter estimates are underestimated, which leads to false-positive results unless a robust variance estimator is used (Kraft et al. 2005).

Although these single locus bi-allelic analyses provide insights, biological inference is limited. For example, the models discussed in the previous paragraphs assume that there are joint maternal–offspring genotype effects but no paternal genotype effects. With the same data, we could have equally plausibly tested for joint paternal–offspring genotype effects or main effects of maternal, paternal, and offspring genotypes. In fact there can be multiple mathematically equivalent parameterizations that have different biological interpretations. Although null hypotheses may be rejected, the statistical analyses cannot provide insights into which biological interpretation is the correct one. Thus it is important, when using these models for gene discovery, not to take the results of any parameterization too literally and recognize a number of alternative, equally plausible explanations may exist (see Sinsheimer et al. 2003 and Ainsworth et al. 2011 for details).

11.3.3 Approaches Using Phenotype, Genetic, and Epigenetic Data

Currently epigenetic data are scarce in epidemiological studies, particularly at the genome-wide level. The most commonly available genome-wide epigenetic data are DNA methylation profiles (Cortessis et al. 2012). Studies collecting these methylation profiles are still small in scale, chiefly because of the expense. The majority of studies use samples from unrelated individuals. Studies of relatives have mainly consisted of twin studies (Bell and Spector 2012; Bocklandt et al. 2011). The predominant use of twin studies is due to (1) the strong tradition of using twins in heritability studies, which provides a wealth of readily available analysis tools; (2) the expense of using full pedigree data; and (3) changes in DNA methylation profile over the course of an individual's lifetime making comparisons of relatives across generations more complicated than using twins.

The data are often expressed as the fraction of a specific CpG site that is methylated (see Laird 2010 for a review of technologies). In statistical analyses,

this fraction, called the beta value, is sometimes treated as an outcome (the ultimate phenotype of interest), and sometimes treated as an intermediate phenotype associated with an outcome. When treating the beta value as an outcome, all the existing quantitative trait analysis approaches, both for data from unrelated individuals or related individuals, can be used including penalized regression (Bocklandt et al. 2011). The heritability of beta values can be calculated by using pedigrees as well as by using twins (Bjornsson et al. 2008). One potential complication with pedigree data is the strong age dependence of the beta values at many DNA methylation sites (Bjornsson et al. 2008; Bocklandt et al. 2011); however in analogy to age dependence for clinical outcomes age can be included as a covariate (e.g., Watanabe et al. 1999; Kangas-Kontio et al. 2010).

A beta value can also be an intermediate phenotype (like a biomarker) of an outcome. Again there are statistical genetic methods that can use unrelated or related individuals and treat beta values as intermediate phenotypes in association studies (Cortessis et al. 2012). One question following from these association studies is: are these epigenetic changes causal or are they responses to the clinical phenotype? Statistical approaches for inferring causality such as Mendelian randomization (Thomas and Conti 2004), genetical genomics (Li et al. 2005), and structural equation modeling (Morris et al. 2010) provide frameworks for answering this question. Using genetic loci associated with the beta values can anchor the causal direction. These three statistical approaches are somewhat related and for space considerations, we focus on Mendelian randomization as it has been used most frequently for the epigenetic explorations.

In the epigenetic context, Mendelian randomization resolves the question of directionality between beta value and an outcome by examining the effect of introducing a genetic covariate, a proxy, into the analysis (Thomas and Conti 2004). The assumption is that this proxy is directly related to the beta value, but it is only indirectly related to the outcome. Thus the magnitude of the true causal effect of methylation at the CpG site on the outcome is the ratio of the magnitude of effect of the genotype on the outcome divided by the magnitude of the effect of genotype on the beta value.

Because other measured covariates such as age, sex, body mass index, or specific biomarkers like lipid levels can be associated with both the beta value and outcome, it may be hard to discern causality. For example, suppose there is an association of age with the clinical phenotype, and there is an association of age with the beta value at a specific CpG site. Is the age effect for the phenotype manifested through DNA methylation? One promising approach to answering this question is two-step Mendelian randomization (Relton and Davey Smith 2012). In this context, the biomarkers, age, etc. constitute exposures. In the first step, a genetic proxy associated with the exposure is used to determine the causality of the exposure for the beta value. In the second step, a different genetic proxy, independent of the first proxy and associated with the beta value, is used to determine the causality of DNA methylation at the CpG site for the outcome.

Of particular relevance to understanding transgenerational effects is that Mendelian randomization can be applied to family data (e.g., Morris et al. 2009). Two-step Mendelian randomization can also span generations. Relton and Davey Smith (2012) discuss the example of maternal alcohol use during pregnancy as the exposure, offspring methylation fraction at a particular CpG site as the intermediate phenotype and offspring cognition as the outcome. In this case, an appropriate genetic proxy for alcohol consumption is the mother’s genotype at an associated locus and an appropriate genetic proxy for the beta value is the offspring’s genotype at another locus, unlinked and independent of the first locus.

11.4 Discussion

Epidemiological study designs and statistical genetic approaches make it possible to detect transgenerational effects in humans. Table 11.2 summarizes the study samples presented and the nature of transgenerational effects that can be determined using them. Determining the correct form of the transgenerational effects using the epidemiological studies is difficult but the more genetic and epigenetic information available, the better the chances of differentiating between the possibilities. Researchers need to be mindful that even detailed epigenetic data are of limited value if the study design is inadequate. The model complexity cannot exceed what is possible given the study sample. For example, none of these transgenerational effects can be tested if the study sample is limited to unrelated cases and controls.

Table 11.2 Examples of study samples and research questions

Study sample	Genetic effect						Prenatal, postnatal, maternal inherited
	Offspring	Maternal	Paternal	Parent of origin	MFG	Long-range transgenerational	
Case–control	Yes	No	No	No	No	No	No
CMCM	Yes	Yes	No	No	No	No	No
CFCF	Yes	No	Yes	No	No	No	No
CPT	Yes	Yes	Yes	Yes	Yes	No	No
Nuclear families	Yes	Yes	Yes	Yes	Yes	No	No
Extended Pedigrees	Yes	Yes	Yes	Yes	Yes	Yes	Yes ^a
Families using ART	Yes	Yes	Yes	Yes	Yes	Yes	Yes

CMCM case-mother, control-mother study sample, CFCF case-father, control-father study sample, CPT case-parent trio study sample

^aYes, if adopted offspring and offspring of sisters are included

Researchers should remember when analyzing data under particular hypotheses that more than one parameterization with different biological interpretations are mathematically equivalent or nearly equivalent. They should also remember that violation of the underlying (and sometime unstated) modeling assumptions may lead to rejection of the null hypothesis without the alternative hypothesis actually being true. For example, violation of the symmetric mating assumption will lead to false inference of maternal effects when analyzing genotype data with CPTs (Sinsheimer et al. 2003). When possible, these modeling assumptions should be checked. Independent mechanistic data from functional studies, in vitro or using model organisms, will be needed to move beyond these associations and resolve these alternative explanations. Despite these caveats, epidemiological data still provide us with strong evidence in support of the existence of transgenerational genetic effects in humans and their roles in complex disease. Moreover they generate hypotheses for further research into the mechanisms of these transgenerational effects.

Acknowledgments We thank the editors for their comments on this chapter. This review was supported in part by NIH grants GM53275 and T32-HG002536.

References

- Ainsworth HF, Unwin J, Jamison DL, Cordell HJ (2011) Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genet Epidemiol* 35:19–45
- Bell JT, Spector TD (2012) DNA methylation studies using twins: what are they telling us? *Genome Biol* 13:172
- Benyshek DC, Martin JF, Johnston CS (2001) A reconsideration of the origins of the type 2 diabetes epidemic among Native Americans and the implications for intervention policy. *Med Anthropol* 20:25–64
- Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrum TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299:2877–2883
- Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S, Vilain E (2011) Epigenetic predictor of age. *PLoS One* 6:e14821
- Cannon M, Jones P, Murray R (2002) Obstetric complications and schizophrenia: historical and meta-analytic review. *Am J Psychiatry* 159:1080–1092
- Chen J, Zheng H, Wilson ML (2009) Likelihood ratio tests for maternal and fetal genetic effects on obstetric complications. *Genet Epidemiol* 33:526–538
- Childs EJ, Palmer CG, Lange K, Sinsheimer JS (2010) Modeling maternal-offspring gene-gene interactions: the extended MFG test. *Genet Epidemiol* 34:512–521
- Childs EJ, Sobel EM, Palmer CG, Sinsheimer JS (2011) Detection of intergenerational genetic effects with applications to HLA-B matching as a risk factor for schizophrenia. *Hum Hered* 72:161–172
- Cordell HJ, Barratt BJ, Clayton DG (2004) Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 26:167–185

- Cordell HJ (2004) Properties of case/pseudocontrol analysis for genetic association studies: effects of recombination, ascertainment, and multiple affected offspring. *Genet Epidemiol* 26:186–205
- Cortessis VK, Thomas DC, Levine AJ, Breton CV, Mack TM, Siegmund KD, Haile RW, Laird PW (2012) Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. *Hum Genet* 131:1565–1589
- Dahlquist GG, Patterson G, Soltesz G (1999) Perinatal risk factors for childhood type 1 diabetes in Europe. The EURODIAB substudy 2 study group. *Diabetes Care* 22:1698–1702
- Edwards TL, Gao X (2012) Methods for detecting and correcting for population stratification. *Curr Protoc Hum Genet Chap1*:1.22.1–14
- Gluckman PD, Hanson MA, Beedle AS (2007) Non-Genomic transgenerational inheritance of disease risk. *BioEssays* 29:145–154
- Gorlova OY, Lei L, Zhu D, Weng SF, Shete S, Zhang Y, Li WD, Price RA, Amos CI (2007) Imprinting detection by extending a regression-based QTL analysis method. *Hum Genet* 122:159–174
- Guilmatre A, Sharp AJ (2012) Parent of origin effects. *Clin Genet* 81:201–209
- Guyton AC (1981) Textbook of medical physiology. W. B. Saunders, Philadelphia, PA
- Harney S, Newton J, Milicic A, Brown MA, Wordsworth BP (2003) Non-inherited maternal HLA alleles are associated with rheumatoid arthritis. *Rheumatology* 42:171–174
- Hollister J, Laing P, Mednick S (1996) Rhesus incompatibility as a risk factor for schizophrenia in male adults. *Arch Gen Psychiatry* 53:19–24
- Hsieh H, Palmer CGS, Sinsheimer JS (2006a) Allowing for missing data at highly polymorphic genes when testing for maternal, offspring and maternal–fetal genotype incompatibility effects. *Hum Hered* 62:165–174
- Hsieh HJ, Palmer CG, Harney S, Chen HW, Bauman L, Brown MA, Sinsheimer JS (2007) Using the maternal–fetal genotype incompatibility test to assess non-inherited maternal HLA-DRB1 antigen coding alleles as rheumatoid arthritis risk factors. *BMC Proc* 1:S124
- Hsieh HJ, Palmer CG, Harney S, Newton JL, Wordsworth P, Brown MA, Sinsheimer JS (2006b) The v-MFG test: investigating maternal, offspring and maternal–fetal genetic incompatibility effects on disease and viability. *Genet Epidemiol* 30:333–347
- Insel BJ, Brown AS, Bresnahan MA, Schaefer CA, Susser ES (2005) Maternal–fetal blood incompatibility and the risk of schizophrenia in offspring. *Schizophr Res* 80:331–342
- Juul-Dam N, Townsend J, Courchesne E (2001) Prenatal, perinatal, and neonatal factors in autism, pervasive developmental disorder-not otherwise specified, and the general population. *Pediatrics* 107:e63
- Kangas-Kontio T, Kakko S, Tamminen M, von Rohr P, Hoeschele I, Juvonen T, Kere J, Savolainen MJ (2010) Genome scan for loci regulating HDL cholesterol levels in Finnish extended pedigrees with early coronary heart disease. *Eur J Hum Genet* 18:604–613
- Khoury MJ, Beaty TH, Cohen BH (1993) Fundamentals of genetic epidemiology. Oxford University Press, Oxford
- Kistner EO, Weinberg CR (2004) Method for using complete and incomplete trios to identify genes related to a quantitative trait. *Genet Epidemiol* 27:33–42
- Klip H, Verloop J, van Gool JD, Koster ME, Burger CW, van Leeuwen FE, OMEGA Project Group (2002) Hypospadias in sons of women exposed to diethylstilbestrol in utero: a cohort study. *Lancet* 35:1102–1107
- Kraft P, Hsieh H-J, Cordell HJ, Sinsheimer J (2005) A conditional-on-exchangeable parental-genotypes likelihood that remains unbiased at the causal locus under multiple-affected-sibling ascertainment. *Genet Epidemiol* 29:87–90
- Kraft P, Palmer CGS, Woodward AJ, Turunen JA, Minassian S, Paunio T, Lonnqvist J, Peltonen L, Sinsheimer JS (2004) *RHD* maternal–fetal genotype incompatibility and schizophrenia: extending the MFG test to include multiple siblings and birth order. *Eur J Hum Gen* 12:192–198

- Laird NM, Lange C (2010) The fundamentals of modern statistical genetics. Springer Verlag, New York, NY
- Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11:191–203
- Lange K (2002) Mathematical and statistical methods for genetic analysis, 2nd edn. Springer Verlag, New York, NY
- Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM (2013) Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics* 29:1568–1570
- Li H, Lu L, Manly KF, Chesler EJ, Bao L, Wang J, Zhou M, Williams RW, Cui Y (2005) Inferring gene transcriptional modulator relations: a genetical genomics approach. *Hum Mol Genet* 14:1119–1125
- Minassian SL, Palmer CGS, Turunen JA, Paunio T, Lonnqvist J, Peltonen L, Woodward AJ, Sinsheimer JS (2006) Incorporating serotypes into family based association studies using the MFG test. *Ann Human Gen* 70:541–553
- Morris NJ, Gray-McGuire C, Stein CM (2009) Mendelian randomization in family data. *BMC Proc* 15:S7, S45
- Morris NJ, Elston RC, Stein CM (2010) A framework for structural equation models (SEM) in general pedigrees. *Hum Hered* 70:278–286
- Nadeau JH (2009) Transgenerational genetic effects on phenotypic variation and disease risk. *Hum Mol Genet* 18:R202–R210
- Ott J (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 26:588–597
- Palmer CGS, Hsieh H, Reed EF, Lonnqvist J, Peltonen L, Woodward JA, Sinsheimer JS (2006) HLA-B maternal–fetal genotype matching increases risk of schizophrenia. *Am J Hum Genet* 79:710–715
- Palmer CGS, Turunen JA, Sinsheimer JS, Minassian S, Paunio T, Lonnqvist J, Peltonen L, Woodward JA (2002) *RHD* maternal–fetal genotype incompatibility increases schizophrenia susceptibility. *Am J Hum Genet* 71:1312–1319
- Relton CL, Davey Smith G (2012) Two-step epigenetic Mendel randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 41:161–176
- Robson EB (1955) Birth weight in cousins. *Annals Hum Genet* 262–8
- Sham PC, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 59:323–336
- Sinsheimer JS, Blangero J, Lange K (2000) Gamete-competition models. *Am J Hum Genet* 66:1168–1172
- Sinsheimer JS, Palmer CGS, Woodward AJ (2003) The maternal–fetal genotype incompatibility test: detecting genotype combinations that increase risk for disease. *Genet Epidemiol* 24:1–13
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Stratchen T, Reed A (2003) Human molecular genetics, 3rd edn. Garland Science/Taylor & Francis Group, Oxford
- Stubbs EG, Ritvo ER, Mason-Brothers A (1985) Autism and shared parental HLA antigens. *J Am Acad Child Psych* 24:182–185
- Terwilliger JD, Ott J (1992) A haplotype based ‘haplotype relative risk’ approach to detecting allelic associations. *Hum Hered* 42:337–346
- Thapar A, Harold G, Rice F, Ge X, Boivin J, Hay D, van den Bree M, Lewis A (2007) Do intrauterine or genetic influences explain the foetal origins of chronic disease? A novel experimental method for disentangling effects. *BMC Med Res Methodol* 7:25
- Thomas DC (2004) Statistical methods in genetic epidemiology. Oxford University Press, Oxford
- Thomas DC, Conti DV (2004) Commentary: the concept of ‘Mendelian Randomization’. *Int J Epidemiol* 33:21–25

- Vermeulen SH, Shi M, Weinberg CR, Umbach DM (2009) A hybrid design: case-parent triads supplemented by control-mother dyads. *Genet Epidemiol* 33:136–144
- Watanabe RM, Valle T, Hauser ER, Ghosh S, Eriksson J, Kohtamèki K, Ehnholm C, Tuomilehto J, Collins FS, Bergman RN, Boehnke M (1999) Familiality of quantitative metabolic traits in Finnish families with non-insulin-dependent diabetes mellitus. Finland-United States Investigation of NIDDM Genetics (FUSION) Study investigators. *Hum Hered* 49:159–168
- Wheeler E, Cordell HJ (2007) Quantitative trait association in parent offspring trios: extension of case/pseudocontrol method and comparison of prospective and retrospective approaches. *Genet Epidemiol* 31:813–833
- Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62:969–978
- Zhou JJ, Pelka S, Lange K, Palmer CG, Sinsheimer JS (2011) Dissecting prenatal, postnatal and inherited effects: ART and design. *Genet Epidemiol* 35:437–446

Chapter 12

Transmission Ratio Distortion: A Neglected Phenomenon with Many Consequences in Genetic Analysis and Population Genetics

Aurélie Labbe, Lam Opal Huang, and Claire Infante-Rivard

Abstract Transmission ratio distortion (TRD) is defined as a statistical departure from the Mendelian 1:1 inheritance ratio and occurs when one of the two alleles from either parent is preferentially transmitted to the offspring (Pardo-Manuel de Villena and Sapienza 2001). This phenomenon is conventionally assessed by the transmission disequilibrium test (TDT) (Spielman et al. 1993), which measures the departure from the expected transmission of an allele from heterozygous parents to affected offspring. In such cases, a departure from Mendelian ratios suggests the presence of linkage and association between the allele and the offspring condition. The TDT and other family-based tests of transmission for linkage disequilibrium and association have been used extensively as one way to provide validation for case–control results while controlling for population structure bias. However, TRD has also been empirically observed in offspring unselected for disease (Infante-Rivard and Weinberg 2005; Naumova et al. 1998; Paterson et al. 2003, 2009; Zollner et al. 2004), which suggests the occurrence of the TRD phenomena in apparently unaffected populations. Although its extent in the human genome is not yet well known, it has also been extensively identified in other species such as mice (LeMaire-Adkins and Hunt 2000; Lyon 2003; Wu et al. 2005), drosophila (Novitski 1951; Sturtevant 1936; Zimmering 1955), and lesser kestrel (Aparicio et al. 2010). Many of the reported TRD loci play a role in tumor suppression and have been found in colon cancer, leukemia, bladder cancer, intestinal adenoma, node-positive breast cancer, and other cancers (De Rango et al. 2007; Eaves et al. 1999; Naumova

A. Labbe (✉)

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada H3A 1A3

Douglas Mental Health University Institute, Montreal, QC, Canada H4H 1R3

e-mail: aurelie.labbe@mcgill.ca

L.O. Huang • C. Infante-Rivard

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, QC, Canada H3A 1A3

et al. 2001; Paterson et al. 2009). A number of TRD loci are within gene regions responsible for imprinting (Eversley et al. 2010; Naumova et al. 2001; Yang et al. 2008), such as D12Nds2 on chromosome 12 and H19 on 11p15.5, leading to loss of imprint and embryonic lethality. Many TRD loci have also been linked to abnormal development in neurogenesis, neuronal differentiation, and other cognitive functions in the central and peripheral nervous system (De Rango et al. 2007; Eversley et al. 2010; Naumova et al. 2001; Paterson et al. 2009; Paterson and Petronis 1999; Riess et al. 1997).

12.1 Introduction

Transmission ratio distortion (TRD) is defined as a statistical departure from the Mendelian 1:1 inheritance ratio and occurs when one of the two alleles from either parent is preferentially transmitted to the offspring (Pardo-Manuel de Villena and Sapienza 2001). This phenomenon is conventionally assessed by the transmission disequilibrium test (TDT) (Spielman et al. 1993), which measures the departure from the expected transmission of an allele from heterozygous parents to affected offspring. In such cases, a departure from Mendelian ratios suggests the presence of linkage and association between the allele and the offspring condition. The TDT and other family-based tests of transmission for linkage disequilibrium and association have been used extensively as one way to provide validation for case-control results while controlling for population structure bias. However, TRD has also been empirically observed in offspring unselected for disease (Infante-Rivard and Weinberg 2005; Naumova et al. 1998; Paterson et al. 2003, 2009; Zollner et al. 2004), which suggests the occurrence of the TRD phenomena in apparently unaffected populations. Although its extent in the human genome is not yet well known, it has also been extensively identified in other species such as mice (LeMaire-Adkins and Hunt 2000; Lyon 2003; Wu et al. 2005), drosophila (Novitski 1951; Sturtevant 1936; Zimmering 1955), and lesser kestrel (Aparicio et al. 2010). Many of the reported TRD loci play a role in tumor suppression and have been found in colon cancer, leukemia, bladder cancer, intestinal adenoma, node-positive breast cancer, and other cancers (De Rango et al. 2007; Eaves et al. 1999; Naumova et al. 2001; Paterson et al. 2009). A number of TRD loci are within gene regions responsible for imprinting (Eversley et al. 2010; Naumova et al. 2001; Yang et al. 2008), such as D12Nds2 on chromosome 12 and H19 on 11p15.5, leading to loss of imprint and embryonic lethality. Many TRD loci have also been linked to abnormal development in neurogenesis, neuronal differentiation, and other cognitive functions in the central and peripheral nervous system (De Rango et al. 2007; Eversley et al. 2010; Naumova et al. 2001; Paterson et al. 2009; Paterson and Petronis 1999; Riess et al. 1997).

TRD mechanisms are not all well understood yet. They include germline selection during mitosis of germ cells, meiotic drive during female meiosis, gametic competition of sperm to achieve fertilization, and embryo lethality due to

deleterious genotype or mother–fetal incompatibility. Furthermore, some epigenetic mechanisms underlying genomic imprinting have also been identified such as imprint resetting error or faulty imprint maintenance at fertilization or in early embryonic development stage.

Since TRD involves a deviation from the Mendelian 1:1 ratio of allelic transmission from parents to offspring, it can only be measured in family-based studies. However, the presence of TRD in populations unselected for disease has a strong impact on the interpretation of results from family-based linkage and association studies designed to detect departure from the expected in sharing or transmission of marker alleles, respectively (Greenwood and Morgan 2000; Paterson et al. 2003, 2009; Zollner et al. 2004). When TRD exists in a population unselected for disease, a linkage or association signal with the TRD locus would be detected in case samples even when there is no linkage or association present. By inflating or attenuating the linkage or association signals, the presence of TRD can therefore lead to false-positive or false-negative allele-sharing or TDT-like test results and induce significant power loss (Greenwood and Morgan 2000). These aspects have not been sufficiently emphasized in the literature and will be addressed in this chapter.

The TRD phenomenon also has implications for developmental genetics. When TRD repeatedly occurs over many generations, the frequency of the allele favored in the selection, and of the alleles at nearby loci, begin to shift upwards in the population (Chevin and Hospital 2006); as a consequence, the disadvantaged allele at the TRD locus gradually becomes rare in the population. Although the impact of TRD in the search for disease-associated rare variants has not, to our knowledge, been investigated in the literature yet, this chapter addresses the issue of the link between TRD loci and rare variants. With the advent of high throughput sequencing of whole-exomes or whole-genomes, the focus is now on rare disease-causing variants. However, since such variants are likely to be seen only once or twice in samples of thousands of unrelated individuals, family studies are enjoying a resurgence in popularity. Indeed, from the basic principle of inheritance, rare disease-causing variants are likely to be seen in multiple members of a family with a high prevalence of the disease. In this context, studying the impact of TRD on the identification of rare variants in family-based studies is very relevant and may provide insights into the interpretation of family-based association study results.

This chapter is divided into six sections. In the first section, a review of study designs and statistical methods to detect TRD is presented, with a particular emphasis on the underlying biological mechanism. In the second and third sections, we revisit the TRD phenomenon, by considering it as a confounding signal in linkage or association studies. Section 12.4 includes a simulation study underlying the importance of control samples to detect and separate TRD signal from association or linkage signal in affected offspring samples. The fifth section addresses TRD from a population genetics perspective and presents the results of a simulation study investigating the link between TRD and rare variants. Finally, Sect. 12.6 presents a case study on thrombophilic gene variants showing how TRD can mask the

association between these variants and the outcome of small-for-gestational-age babies, and how by obtaining TRD estimates from control samples one can recover association signals that would be left undetected otherwise.

12.2 TRD Inference: Study Designs and Methods

TRD is the result of disruptive mechanisms during the gametic or embryonic development stages (see Huang (2013) for a review of such mechanisms). These TRD mechanisms lead to differential survival in embryos and include germline selection during mitosis of germ cells (Hastings 1991), meiotic drive during female meiosis (Pardo-Manuel de Villena and Sapienza 2001), gametic competition of sperm to achieve fertilization (Zollner et al. 2004), embryo lethality due to deleterious genotype or mother–fetal incompatibility (Zollner et al. 2004), as well as imprint resetting error or faulty imprint maintenance at fertilization or in early embryonic development stage (Naumova et al. 1995, 2001; Yang et al. 2008).

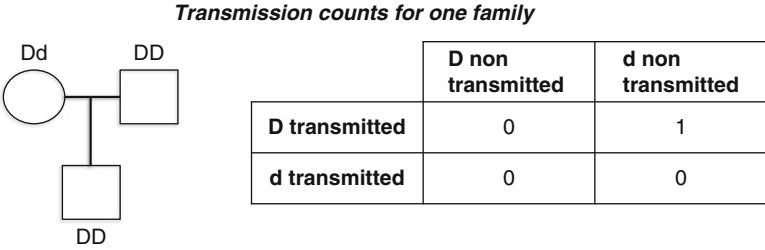
If the search for TRD loci is unrelated to a specific disease but rather the primary research goal, families with offspring unselected for phenotype or disease should be genotyped. Depending on the underlying biological mechanism, TRD can be observed in different family structures ranging from two-generation families to larger multigenerational families. These different scenarios are reviewed in detail in the following sections.

12.2.1 Detecting TRD Using a Transmission Disequilibrium Test (TDT) Approach in Trios with Offspring Unselected for Phenotype

Departure from the expected transmission probabilities of an allele from heterozygous parents to offspring is conventionally measured with the TDT in a sample of trios (parents and their offspring) (Spielman et al. 1993). Consider for example a TRD locus with 2 alleles, D and d, where the allelic transmission ratio from parent to unaffected offspring is $D:d = k:1$ (i.e., the D allele is transmitted k times more often than the d allele). Assuming two heterozygote parents with Dd genotype, the expected proportion of offspring genotypes is given in Table 12.1.

Table 12.1 Distribution of offspring genotype proportions for different values of TRD ratio

TRD ratio	Offspring genotype		
	DD	Dd	dd
$k = 1$ (Mendelian transmission)	$1/4 = 0.25$	$1/2 = 0.5$	$1/4 = 0.25$
$k = 1.5$ (TRD with ratio $D:d = 1.5:1$)	$9/25 = 0.36$	$12/25 = 0.48$	$4/25 = 0.16$
$k = 2$ (TRD with ratio $D:d = 2:1$)	$4/9 = 0.44$	$4/9 = 0.44$	$1/9 = 0.11$
$k = 3$ (TRD with ratio $D:d = 3:1$)	$9/16 = 0.56$	$6/16 = 0.37$	$1/16 = 0.062$



TDT in a sample of n=90 families with heterozygote parents

	D non transmitted	d non transmitted	Total
D transmitted	a = 0	b = 120	120
d transmitted	c = 60	d = 0	60
Total	60	120	2n=180

The TDT tests for the null hypothesis of Mendelian allelic transmission D:d=1:1

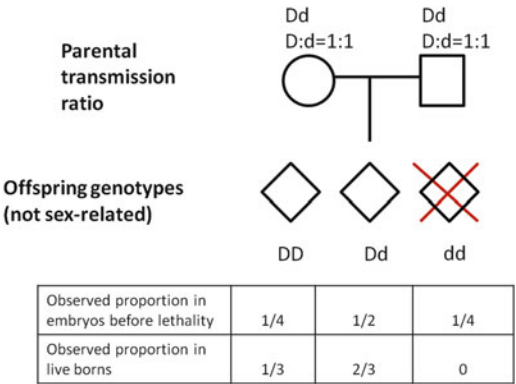
Null hypothesis $H_0: \frac{b}{b+c} = \frac{c}{b+c} = 0.5 \quad \chi^2 = \frac{(b-c)^2}{b+c} = 20 \quad \text{P-value is } 7.7 \times 10^{-6}$

Fig. 12.1 Illustration of the TDT

In this example, observed offspring genotypes do not obey the Mendelian ratio when $k > 1$, leading to a departure from the expected distribution. This form of TRD can be explained by a collection of biological mechanisms referred to as germline selection occurring during mitosis or by embryo lethality. Germline selection refers to mechanisms such as mutation, recombination, and gene conversion, which cause cells with certain genotypes to be produced at a higher proportion than others. Hence, germ cells entering the meiosis stage have an imbalanced genotype ratio. When trio samples are collected, the over-transmission of a marker allele from heterozygote mothers or fathers is tested using the TDT, which is essentially a McNemar statistical test (Liddell 1976). This process is illustrated in Fig. 12.1.

Over-transmission of a marker allele from parents to offspring can also occur in a sex-of-parent-specific manner, which can be explained by female meiotic drive or by gametic competition. Meiotic drive occurs when a haplotype with structural advantage tends to be transmitted more during meiosis. Gametic competition (also called gametic selection) refers to the competition of sperms surviving through meiotic drive to achieve fertilization. In principle, these TRD mechanisms can be uncovered using the TDT with trios where over-transmission within strata of heterozygote mothers or heterozygote fathers is tested using a McNemar test. However, when both parents are heterozygous, TDT on mothers versus TDT on fathers is no longer a valid test due to lack of statistical independence of transmissions (Weinberg 1999). Other tests have been proposed in determining parent-of-origin effect, such as Transmission Asymmetry Test (TAT) (Weinberg et al. 1998), Likelihood Ratio Test (LRT) (Weinberg 1999), and Parental

Fig. 12.2 TRD caused by embryo lethality. We assume here that the mutant allele is *d* and that lethality is autosomal recessive. As a result, *dd* genotype is eliminated before birth



Asymmetry Test (PAT) (Weinberg 1999; Zhou et al. 2009). However, these tests require the absence of prenatal maternally mediated effect, defined as the effect of maternal genotype on outcome. This requirement is justified by the fact that maternal effect can cause differential weighting of the maternal and paternal transmissions. For the scenario where diseases are subject to prenatal maternally mediated effects, the Parent-of-Origin Likelihood Ratio Test (PO-LRT) method remains the only valid testing procedure (Weinberg 1999).

TRD can also be caused by other mechanisms of selection not occurring in the parents but after the embryo is formed. Such mechanisms, termed embryo lethality, occur when embryos with a specific genotype are eliminated. Because this leads to an imbalance in the offspring genotypic ratios as illustrated in Fig. 12.2, a TDT approach can also be used.

Embryo lethality can also be sex-specific, which induces a sex-of-offspring-specific TRD. The analytical strategy is the same as above, except that TDT is performed only in female (respectively male) offspring. Note that the issues related to maternally mediated effects discussed above are also relevant in this context.

Unfortunately, since TDT looks at over-transmission of a marker allele where embryos with the faulty genotype could not have survived, it is impossible to determine whether TRD was caused by mechanisms occurring in the parents (meiotic drive, gametic competition, or other mechanisms) or at the embryonic stage (embryo lethality). Note also that in the following developments, the TDT approach can be applied beyond trios to extended families unselected for phenotype (Tiwari et al. 2008).

12.2.2 Detecting TRD in Extended Families Unselected for Phenotype Using Nonparametric Linkage Analysis

In the case of extended families, nonparametric linkage analysis can be used as an alternative to the TDT approach. Nonparametric analysis looks at over-sharing of alleles identical by descent (IBD) between “affected” related pairs. Two or more

alleles are said to be IBD if they are identical copies of the same ancestral allele. An over-sharing of alleles IBD between related affected individuals at a specific marker indicates linkage between this marker and the disease susceptibility locus. In order to identify TRD in families unselected for disease, all offspring are considered as “affected”, which essentially means “having survived”. Therefore, the objective is to determine regions in the genome linked to the phenotype defined as “being alive in the last generation” (Paterson et al. 2009). This analytical strategy was used by Paterson et al. (2009) in the Framingham Heart Study cohort, but no loci met the genome-wide criteria for linkage. Note that the case of sex-specific TRD can also be investigated by performing linkage analysis separately in males or females.

As discussed before, when studying over-sharing of a marker allele where embryos with the faulty genotype could not have survived, it is impossible to determine whether the observed TRD occurred in the parents or at the embryonic stage. As a result, the underlying biological mechanisms driving TRD such as germline selection, meiotic drive, gametic competition or embryo lethality cannot be differentiated and therefore identified precisely.

12.2.3 Grandparental Origin TRD: Imprinting Errors

In the types of TRD described above, deviation in the allelic transmission from the Mendelian ratio is inferred based on what is observed in the offspring genotypes. Another form of TRD can occur which is induced by an imbalance in the grandparental origin of the offspring’s genotypes. Under Mendelian inheritance in humans, each individual contains the genetic information transmitted by his/her four grandparents, with an expected transmission ratio of 1:1:1:1. However, a deviation from this ratio, which is also a form of TRD, can be explained by possible imprint resetting errors in the parent’s germline or by erroneous maintenance of parental imprints in early embryonic development stage. Figure 12.3 illustrates an example of a three-generation family with correct imprint resetting and maintenance. In this example, we assume that the genetic locus is maternally imprinted, which means that only paternal alleles are expressed in offspring. As we see in Fig. 12.3, imprint marks have been correctly reset in grandparents A, B, C and D, so that each egg cell contains a maternal imprint and each sperm cell contains a paternal imprint. As a result, both individuals in the second generation inherit a correctly imprinted allele from their mother and a correctly non-imprinted allele from their father. The same resetting process successfully occurs in the germline of the second generation individuals (father and mother) before meiosis. Then, when the egg from the mother is fertilized by the sperm of the father, each of them transmits a correctly imprinted allele to the offspring. As seen in Fig. 12.3, there is no deviation from the Mendelian ratio in either the offspring genotypic ratios, nor in the allelic origin of parents and grandparents.

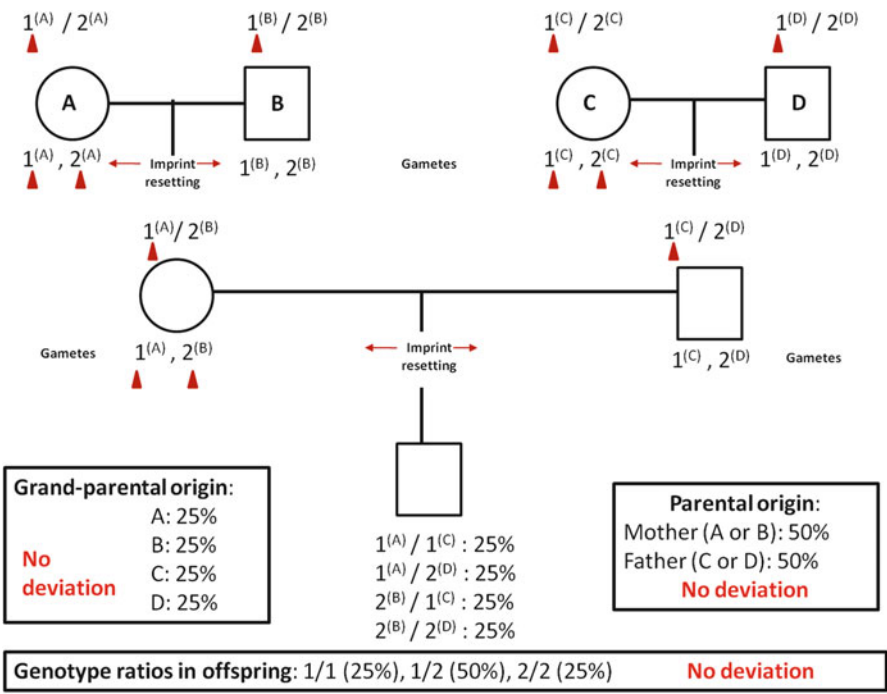


Fig. 12.3 Example of a three-generation family including four grandparents, two parents and offspring. We consider a marker with two alleles, denoted as 1 and 2. Grandparents are denoted as A, B, C and D and superscripts at each genotype indicate the grandparent origin. In this example, correct imprint resetting occurs in the germline before the production of eggs and sperm cells. We assume here that the marker is maternally imprinted and imprinted marks are represented by a red triangle

Figure 12.4 illustrates the scenario where an imprint resetting error occurred on allele 2 of the mother, which is incompatible with embryonic survival. This leads to the deviation from Mendelian inheritance ratio in the allelic origin of the grandparents. Interestingly, this also leads to a deviation from the Mendelian ratio in the offspring, which seems to suggest that this phenomenon could be captured by using the TDT approach in trios described above.

For comparison, Fig. 12.5 illustrates a similar scenario, but the imprint resetting error occurred on allele 1 of the father. Similarly, the allele which failed to reset correctly is under-transmitted. A deviation from Mendelian ratio of the alleles from grandparents can be observed in the offspring. This observation is the basis of the statistical analyses aiming to uncover TRD induced by imprinting errors.

Two analytical strategies have been proposed in the literature to determine the grandparental origin of TRD. First, a simple binomial test can be used by determining if the proportions of grandpaternal alleles and grandmaternal alleles are equal in the offspring's genotypes for a given marker. In practice, TRD is estimated by the proportion of grandmaternal alleles transmitted to the offspring (Naumova et al. 2001; Yang et al. 2008). The method of maximum likelihood (Lange 1997)

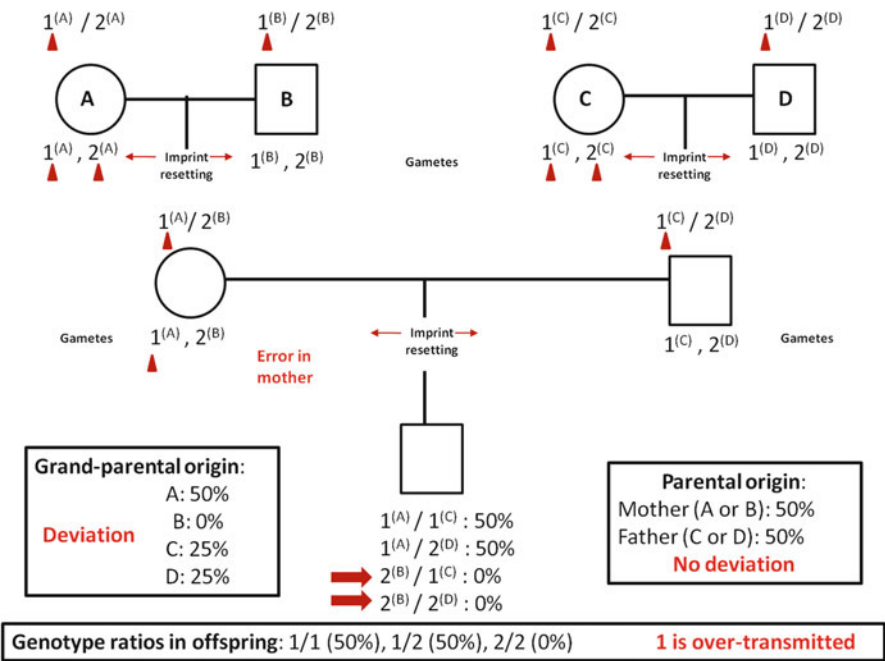


Fig. 12.4 Example of a three-generation family with imprint resetting error at allele 2 in mother. Same scenario as in Fig. 12.3, an imprint resetting error occurred in the mother, which is incompatible for embryonic survival

can be used to estimate TRD in the presence of missing genotypes, by using neighboring flanking markers as well as map distances (Croteau et al. 2002). In cases where embryo lethality due to imprinting error occurs in a sex-of-offspring specific manner, TRD can also be estimated by using a logistic regression model predicting grandparental source (dichotomous outcome), where variables such as sex of offspring and mating type of parents are included in the model (Yang et al. 2008). In Yang et al.'s paper (2008), grandparental origin TRD locus was inferred on the basis of genotypes of the closest microsatellite markers. For non-informative markers, it was inferred on the basis of the grandparental origin of the flanking markers.

12.3 Impact of TRD in Association or Linkage Analysis

When TRD occurs at a disease locus or at a locus in linkage disequilibrium (LD) with the disease locus, a linkage or association signal would be inflated or attenuated, potentially leading to a false positive or false negative result. On the other hand, if TRD occurs at a locus distant from the disease susceptibility locus

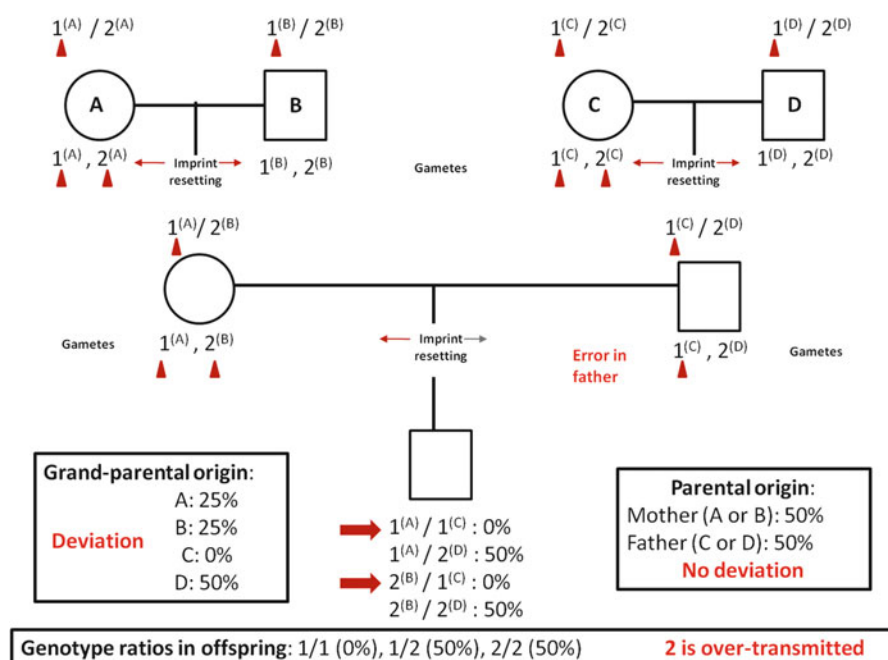


Fig. 12.5 Example of a three-generation family with imprint resetting error at allele 1 in father. Same example as in Fig. 12.3, an imprint resetting error occurred in the father, which is incompatible for embryonic survival

(DSL) and is not in LD with the DSL, a linkage signal at the TRD locus would be detected in families selected for disease, leading to a false positive signal. This phenomenon has been investigated in several studies (Greenwood and Morgan 2000; Paterson et al. 2003, 2009; Zollner et al. 2004) quantifying the impact of TRD on linkage results. Greenwood and Morgan (2000) studied the case of affected sib pairs and showed that IBD patterns between two affected sibs are strongly modified, leading to important bias in the significance of a sib-pair linkage analysis. They also suggested to increase the sample size by a small amount to maintain the desired power if TRD with modest deviation from Mendelian segregation is suspected at the planning stage of a study.

Several approaches have been suggested to overcome the bias induced by TRD in family-based linkage or association analysis. When a TRD locus unrelated to a specific disease is detected in families unselected for phenotype, it is not surprising that most approaches proposing to account for TRD in the statistical analysis use a combination of both non-affected and affected subjects. For instance, Spielman et al. (1993) proposed to use a mixture of case trios (affected offspring with parents) and control trios (unaffected offspring with parents) to differentiate true linkage or association signals from false positives due to TRD by applying a TDT to both types of trios. The study concluded that a statistically significant TDT in case trios but not in control trios suggests evidence of true linkage and association.

Furthermore, a statistically significant TDT observed in both case trios and control trios, with a significant difference in transmission counts between case trios and control trios also suggests evidence of true linkage and association. Note that an extension of TDT using discordant sib-pair-parent tetrads has been proposed by Deng et al. (2009) and can be used to assess significance of the TDT results. Finally, we propose another strategy, which consists in estimating TRD at a marker locus in a sample unselected for disease and modifying the null hypothesis of association accordingly in the sample selected for disease. For example, if the probability to transmit the minor allele from parents to offspring is estimated to be 0.6 in the control sample, one can test the null hypothesis that this probability is 0.6 in the case sample. A deviation from this value should indicate a true signal. This strategy, similar in essence to the one proposed by Spielman (1993) will be detailed and illustrated in the case study section.

In the context of affected sib-pair designs, Lemire et al. (2004) developed a novel allele-sharing statistic that accounts for the possible bias induced by TRD. This statistics evaluates the excess sharing of alleles in pairs of affected siblings, as well as a deficit of sharing in phenotypically discordant relative pairs, where available. If unaffected siblings are available, this statistic is unbiased in TRD regions. If more distantly related unaffected subjects are available, the bias is reduced but not completely eliminated. Note that the Haseman-Elston regression model for affected sib pairs is not affected by TRD and therefore represents another interesting alternative for the analysis.

Casellas (2012) developed a Bayesian binomial model accounting for deviation TRD mechanisms in F2 mouse crosses. This model was used to perform genome-wide scans for TRD quantitative trait loci (QTL) on six such crosses. Results suggest a relevant incidence of TRD phenomena in mouse with important implications for both statistical analyses and biological research, such as those underscored in this chapter.

12.4 The Use of Controls to Differentiate TRD from Real Signal: A Simulation Study

In addressing the phenomenon of TRD, which acts as a confounder for linkage and association signals, Spielman et al. (1993) first suggested the use of both case and control trios. The proposed method is to apply a TDT separately on case trios and control trios. The method is illustrated in Tables 12.2 and 12.3, with marker minor allele denoted as M and major allele as m.

Spielman's study (1993) concluded that:

1. A statistically significant TDT in case trios suggests evidence of either linkage/association, or TRD, or both.
2. A statistically significant TDT in control trios suggests evidence of TRD, or both TRD and linkage/association.

Table 12.2 TDT on case trios and control trios

	Case trios		Control trios	
	Non-transmitted allele		Non-transmitted allele	
Transmitted allele	M	m	M	m
M	a_1	b_1	a_2	b_2
M	c_1	d_1	c_2	d_2
TDT test statistics	$\chi_D^2 = \frac{(b_1 - c_1)^2}{(b_1 + c_1)}$		$\chi_D^2 = \frac{(b_2 - c_2)^2}{(b_2 + c_2)}$	

Table 12.3 Pearson's Chi-square test on case trios and control trios

	Transmitted allele in heterozygous parents		Row total
	M	m	
Case trios	b_1	c_1	n_1
Control trios	b_2	c_2	n_2
Column total	n_b	n_c	n
Pearson's Chi-square test statistic	$\chi_{DC}^2 = \frac{n(b_1c_2 - c_1b_2)^2}{(n_1n_2n_bn_c)}$		

Table 12.4 Simulation results for four scenarios each averaged over 500 simulations based on TDT and Pearson's Chi-square test

Scenarios	Presence of linkage and association	Presence of TRD	Significance of TDT in case trios	Significance of TDT in control trios	Significance of Pearson's Chi-square test of case trios vs. control trios transmission counts
1	No	No	No	No	No
2	No	Yes	Yes	Yes	No
3	Yes	No	Yes	No	Yes
4	Yes	Yes	Yes	Yes	Yes

3. A statistically significant TDT in case trios but not in control trios suggests evidence of true linkage and association.
4. When a statistically significant TDT is observed in both case trios and control trios, a nonsignificant Pearson's Chi-square statistic of case trios versus control trios on transmission counts suggests evidence of TRD only.
5. When a statistically significant TDT is observed in both case trios and control trios, a significant Pearson Chi-square statistic of case trios versus control trios on transmission counts suggests evidence of true linkage/association and TRD.

To verify Spielman et al.'s (1993) findings, we set up a simulation study for the four following scenarios described in Table 12.4.

The disease allele frequency (p) in the population was set between 0.01 and 0.05 indicating a rare to moderately rare disease frequency. The marker minor allele frequency (q) was set at 0.1. The underlying TRD influence on the marker locus had a ratio between 0.6 and 0.9 for the minor allele, exploring mild to extreme skew of transmission; here we define the TRD ratio to be the proportion of the preferred allele transmission counts among all transmission counts from parents to offspring at a specific locus. For example, if it is three times more likely to transmit the advantaged over the disadvantaged allele, the TRD ratio is $3/(3 + 1) = 0.75$. The recombination fraction between disease and marker loci (θ) was specified as 0.1 in the scenarios 3 and 4 when there was linkage and association between disease and marker loci, or otherwise is set to 0.5 (scenarios 1 and 2). A prespecified linkage disequilibrium (LD) parameter (δ) was adjusted for each disease allele frequency being tested, to ensure positive haplotype frequencies, which depend on disease and marker allele frequencies. Therefore, LD was set to be slightly less than the minimum of $p(1 - q)$ and $q(1 - p)$ when there was linkage and association (scenario 3 and 4), and set to 0 otherwise (scenario 1 and 2).

Based on the haplotype frequencies, which depend on the marker and disease allele frequencies, and the LD parameter, we generated a population of 600,000 trios (parents and child). We then simulated random mating in this population. Recombination and transmission of alleles occurred with the probabilities stated above for recombination fraction and TRD ratio. Assuming a recessive mode of inheritance at the disease loci, we randomly sampled 500 case trios and 500 control trios from the simulated population. We then applied the TDT at the marker, for both the case and control trios. As suggested by Spielman et al. (1993), we further applied the Pearson's χ^2 test to assess the excess or deficit in transmission of minor allele over major allele in case trios versus control trios. This procedure was repeated 500 times, and the results of the test statistics were averaged over these 500 simulations. Both the McNemar test and Pearson's χ^2 test are 1 degree of freedom tests, the p -values are computed accordingly using each of the four averaged test statistics over 500 simulations. Our results support the proposals of study design, statistical method, and conclusions suggested by Spielman et al. (1993), as shown in Table 12.4. This simulation study was repeated for a dominant mode of inheritance, and the same results were obtained.

12.5 TRD and Rare Variants: A Simulation Study

The impact of TRD at the organismal level could become manifest at the population level as the human genome evolves over time. Therefore, TRD should also be studied in a population genetics context because such selection leads to changes in the diversity of the population gene pool over generations. Changes in genetic diversity over time culminate in the current population to an equilibrium state of parameters such as the minor allele and haplotype frequencies at TRD and neighboring loci, and linkage disequilibrium between loci. If the TRD selection is

persistent through many generations, a gradual shift in the allele frequency at the TRD locus would be observed. Over time, the positively selected allele(s) could become fixed in the population while the alternatives are completely eliminated. This may provide an explanation as to why studies have been able to discover only a small number of TRD loci, because alleles at some of these TRD loci may have already become monomorphic. Therefore, no genetic variation could be detected in the population on these “disappeared” TRD loci. However, through study of identified known TRD loci, some negatively selected alleles still exist at a low frequency and remain polymorphic as rare variants. This raises questions as to why such a selection process did not sweep the positively selected allele into fixation. Several authors have tried to answer this question by suggesting theories on sources of counter-balancing forces which keep the allele in polymorphic state, such as recombination (Haig and Grafen 1991), mutation and genetic drift (Polaski 1998), and an immunogenetic advantage for survival in later adulthood regardless of low fertility (Westendorp et al. 2001).

The existence of these rare variants provides us with great insight into the understanding of TRD selection and the importance of corresponding gene functions at these loci. Rare disease variants are currently the focus of genome-wide association studies in search of missing heritability in complex disorders (Maher 2008). It has been hypothesized that rare disease variants could be more functional than common variants and have high penetrance (Bodmer and Bonilla 2008; Gorlov et al. 2008; Kryukov et al. 2007). This suggests a potentially similar role for negatively selected TRD rare variants when their gene functions determine survival. Since there is usually low power to detect rare variants using a standard genome-wide genotyping platform with feasible sample size, there are intense ongoing research efforts to address this issue (Cirulli and Goldstein 2010; Li and Leal 2009). These efforts should lead to a better understanding of TRD and its contribution to the rare variant phenomenon itself.

By using the formulae in Chevin and Hospital (Chevin and Hospital 2006), we set up a simulation study to trace the marker allele frequency at a TRD locus over generations. The marker allele frequency (q) at generation 0 is set at 0.1 for the minor allele. Disease allele frequency (p) at a neighboring locus is set at 0.01, which corresponds to a rare disease allele. To simulate presence of linkage and association, the recombination fraction (θ) is specified at 0.1, and LD parameter (δ) at 0.0089, which is slightly less than the minimum of $q(1 - p)$ and $p(1 - q)$ to ensure positive haplotype frequencies. Let r be the TRD ratio and q_i be the marker allele frequency at the i th generation. Here we use a different notation than Chevin and Hospital (2006) to be consistent with our definitions specified above. The change in marker allele frequency in i th generation (q_i), as shown in equation (1) of Chevin and Hospital (2006), is $(2r - 1)q_{i-1}(1 - q_{i-1})$. With rearrangement, the LD in i th generation is approximated by $\delta_i = (1 - 2\theta)q_i(1 - q_i) \delta_{i-1}q_{i-1}(1 - q_{i-1})$, (seen in a different form in equation (9) of Chevin and Hospital (2006)), which decays over time. For an exact computation of LD, the more complex formula in equation (8) of Chevin and Hospital (2006) also includes the TRD ratio on top of the previously mentioned parameters.

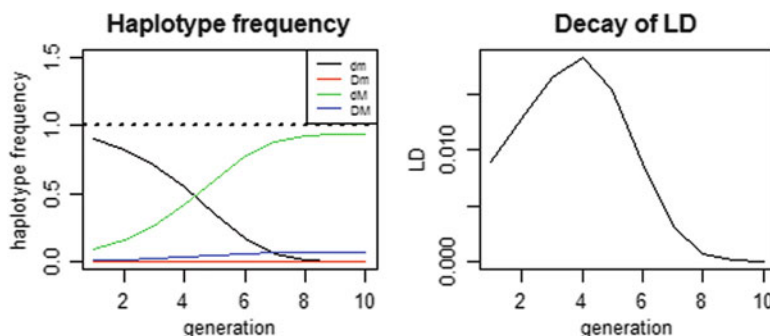


Fig. 12.6 Illustration of genetic diversity over generations when TRD ratio persistently occurs at 0.9, measured by haplotype frequencies of TRD and neighboring loci with minor alleles M and D respectively (*left*), and LD parameter (*right*). At the equilibrium stage, the frequency of allele M at TRD locus is the sum of haplotype frequencies of dM and DM which is $0.935 + 0.065 = 1$, where it has reached fixation

In Fig. 12.6 (*left panel*), we have illustrated the change in diversity in terms of haplotype frequencies of the linked marker and neighboring disease loci with minor alleles M and D respectively, and the LD measure, with a TRD ratio of 0.9. Furthermore, different combinations of TRD ratio, recombination fraction and LD have been experimented to illustrate the range of corresponding number of generations for each combination to reach fixation, as shown in Fig. 12.6 (*right panel*). The figure shows that as recombination fraction decreases or LD parameter increases, it takes longer for the allele at the TRD locus to reach fixation, because the presence of linkage and association slows down the selective sweep. On the other hand, when the TRD ratio decreases, it also takes longer for the allele at the TRD locus to reach fixation because of the milder distortion force.

Through our simulation, we found that for a TRD ratio of 0.9, fixation can be reached in ten generations. As for a TRD ratio of 0.6, it can take up to 80 generations to reach fixation, depending on the strength of linkage and association between marker and disease loci. These changes in genetic diversity over time culminate eventually in an equilibrium state for the involved parameters in the population, namely the minor allele and haplotype frequency at TRD and neighboring loci, and LD between marker and disease loci (Huang et al. 2011).

12.6 TRD and Its Impact on Association Analysis: A Case Study

As mentioned above, a strategy to account for TRD in association analyses consists in estimating TRD at a marker locus in a sample unselected for disease and modifying the null hypothesis of association accordingly in the sample selected

Table 12.5 Transmission counts from a set of trios

	Not transmit M1	Not transmit M2
Transmit M1	a	b
Transmit M2	c	d

for disease. For instance, consider a marker with two alleles, M1 and M2 with frequencies q and $(1 - q)$ respectively. Consider also a disease locus with two alleles and minor allele frequency of p . Let D and θ be the linkage disequilibrium and recombination fraction, respectively between the disease and marker locus. Suppose also that TRD occurs at the marker locus such that the probability to transmit M1 from a parent to the offspring is α . Under these assumptions, one can write the probability that a parent with genotype M1M2 transmits the allele M2 and does not transmit M1:

$$P(\text{Transmit M1, not transmit M2}) = 2\alpha \left(q + \frac{D}{p} \right) (1 - q) - 2\alpha \theta \frac{D}{p}.$$

Similarly, one can compute $P(\text{Transmit M2, not transmit M1})$. Under this setting, testing the null hypothesis H_0 : $P(\text{Transmit M1, not transmit M2}) = \frac{\alpha}{1-\alpha} P(\text{Transmit M2, not transmit M1})$ is a valid test for association since this null hypothesis is equivalent to $\theta = 1/2$ and $D = 0$. If the TRD value α is known from the literature or from an external control sample, one can test for association with the disease locus in a sample of case trios. Assuming transmission counts as in Table 12.5, the chi-square test statistic is equal to
$$\chi^2 = \frac{[(1-\alpha)b - \alpha c]^2}{\alpha(1-\alpha)(b+c)}.$$

Note that this statistics reduces to the conventional TDT statistics $\chi^2 = \frac{(b-c)^2}{(b+c)}$ when $\alpha = 0.5$, which means that no TRD occurs. The advantage of this approach over Spielman’s approach, which compares transmission in control and case samples, is that the TRD value α can be taken from the literature or from an external source, if such information is available.

This approach was applied to a study on the association between thrombophilic gene variants and the occurrence of births qualified as small-for-gestational-age (defined below). Thrombophilia is the name given to a condition characterized by a tendency to develop thrombosis, mostly as a consequence of inherited polymorphisms. The most commonly studied thrombophilic polymorphisms are the *C677T* and *A1298C* variants in the methylene tetrahydrofolate reductase (*MTHFR*) gene; the *G1691A* variant in *Factor V Leiden*; and the *G20210A* variant in the *Prothrombin* (or *Factor II*) gene. Others, less studied are the *plasminogen activator inhibitor-1* (*PAI-1*) gene and the *Factor XIII* (*F XIII*) gene. Small-for-gestational-age (SGA) birth is defined as birth weight below the 10th percentile for gestational age and sex, according to national standards. The underlying hypothesis in our work was that a thrombophilic predisposition will increase the risk of vascular thrombosis, which in turn can lead to placental insufficiency and small infants.

Table 12.6 Results from the TDT tests. *P*-values (*P*) and transmission ratios (α) are presented

	TDT in controls	Modified TDT in cases	Classic TDT in cases
FX111	<i>P</i> = 0.62 α = 0.48	<i>P</i> = 0.43	<i>P</i> = 0.21 α = 0.45
PAI	<i>P</i> = 0.43 α = 0.47	<i>P</i> = 0.61	<i>P</i> = 0.19 α = 0.45
MTHFR-A1298C	<i>P</i> = 0.0016 α = 0.39	<i>P</i> = 0.12	<i>P</i> = 0.09 α = 0.44
MTHFR-C677T	<i>P</i> = 0.28 α = 0.46	<i>P</i> = 0.64	<i>P</i> = 0.11 α = 0.45
FV	<i>P</i> = 0.005 α = 0.27	<i>P</i> = 0.00033	<i>P</i> = 0.44 α = 0.57
FII	<i>P</i> = 0.0002 α = 0.09	<i>P</i> = 0.07	<i>P</i> = 0.002 α = 0.2

The study was described previously (Infante-Rivard et al. 2002, 2003, 2005; Infante-Rivard 2010). It was initiated as a case–control comparison of mother and newborn dyads with and without SGA birth, and we extended the data to form case–parent trios by including genetic material obtained from fathers. All SGA infants seen at our university center in Montréal, Québec, Canada, between May 1998 and June 2000 were eligible if they were singletons who were born alive after the 24th week of gestation without severe congenital anomalies. During that period, 505 case mothers were seen and 493 (98 %) participated in the study. The same criteria applied to the selection of control mothers (defined as women whose babies’ birth weights were at or above the 10th percentile for gestational age and sex). Controls were matched to cases on gestational week, sex, and race/ethnicity and were selected for having a birth date closest in time to the case’s. The mothers of 480 controls were invited to participate, and 472 (98 %) accepted. Medical records were used to determine gestational age on the basis of the obstetrician’s assessment from ultrasound and other clinical data.

Blood was obtained from 448 case newborns (91 % of participants and 89 % of eligible persons) and 431 control newborns (91 % of participants and 90 % of eligible persons). Maternal and umbilical cord blood samples were collected. Approximately midway through the study, we started prospectively collecting DNA samples from fathers. We obtained genetic material on 260 case fathers and 248 control fathers, representing a response rate of approximately 86 % (of those invited to participate). Genotyping is extensively described in previous papers (Infante-Rivard 2010).

First, a TDT was performed in the control sample in order to estimate TRD at the 6 SNPs. Then, a modified TDT as described above was performed in the case sample by adjusting the null hypothesis using the TRD values obtained in the control sample. A classic TDT in cases was also performed for comparison purposes. Results are presented in Table 12.6.

As one can see, a TDT in the control sample leads to the detection of TRD at loci MTHFR-A1298C, FV and FII, with the probability α to transmit the minor allele from heterozygote parent to offspring being 0.39, 0.27 and 0.09 respectively (recall that a probability of 0.5 represents a Mendelian transmission). When testing at each locus the null hypothesis that an α -deviation is also observed in the case sample, we obtained very interesting results. First, we observe that the null hypothesis is strongly rejected at the Factor V locus, which suggests that an extra-deviation from the TRD transmission is present in the case sample. By looking at the transmission counts, we also observed that the probability to transmit the minor allele for heterozygous parent to offspring is 0.57 in the case trios, which does not suggest the presence of association or linkage if a standard TDT test is applied (p -value 0.44). In fact, there are two TRDs acting in opposite directions at this locus. In the control population, a TRD is present which favors the transmission of the major allele. However, in the case population, the minor allele is over-transmitted, which suggests an association with the disease locus. Note that this could not be observed by performing a standard TDT on the case trios since the two acting TRDs would tend to cancel each other and lead to an artificially observed Mendelian transmission in cases.

Most studies on SGA have used a case-control design. Many of the published studies were affected by likely selection bias due to the nature and the way in which controls were selected, leading to spurious results. Nevertheless, despite the fact that not all studies reported an increased risk for FV (Dudding et al. 2004), many studies have done so (Dudding and Attia 2004). The discrepancy between results from case trios alone (often assumed to be a more robust design) and case-control studies may be reconciled with the approach we propose here.

Another interesting locus to study is the Factor II locus. A TRD is observed in the control population as well as in the case population, leading to a significant TDT test in the case trios. However, after adjusting for this observed deviation in the control sample, a TDT test in the case population is no longer significant, which suggest that the association observed in the case trios using a standard TDT was a false positive.

In conclusion, although not well known or appreciated, TRD is a population genetic phenomenon that can provide insight into the evolution of the genome as we find it today; on a more practical level, and as we have argued, considering the possibility of TRD in linkage and association studies is important for the validity of conclusions.

References

- Aparicio JM, Ortego J, Calabuig G, Cordero PJ (2010) Evidence of subtle departures from Mendelian segregation in a wild lesser kestrel (*Falco naumanni*) population. *Heredity* 105:213–219

- Becker T, Jansen S, Tamm S, Wienker TF, Tummler B, Stanke F (2007) Transmission ratio distortion and maternal effects confound the analysis of modulators of cystic fibrosis disease severity on 19q13. *Eur J Hum Genet* 15:774–778
- Blyth K, Vaillant F, Jenkins A, McDonald L, Pringle MA, Huser C, Stein T, Neil J, Cameron ER (2010) Runx2 in normal tissues and cancer cells: a developing story. *Blood Cells Mol Dis* 45:117–123. doi:[10.1016/j.bcmd.2010.05.007](https://doi.org/10.1016/j.bcmd.2010.05.007)
- Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40:695–701
- Botta A, Tacconelli A, Bagni I, Giardina E, Bonifazi E, Pietropolli A, Clementi M, Novelli G (2005) Transmission ratio distortion in the spinal muscular atrophy locus: data from 314 pre-natal tests. *Neurology* 65:1631–1635
- Casellas J, Gualarte RJ, Varona L, Mehrabian M, Schadt EE, Lusk AJ, Attie AD, Yandell BS, Medrano JF (2012) Genome scans for transmission ratio distortion regions in mice. *Genetics* 191(1):247–259
- Chevin LM, Hospital F (2006) The hitchhiking effect of an autosomal meiotic drive gene. *Genetics* 173:1829–1832. doi:[10.1534/genetics.105.052977](https://doi.org/10.1534/genetics.105.052977)
- Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11:415–425
- Croteau S, Andrade MF, Huang F, Greenwood CM, Morgan K, Naumova AK (2002) Inheritance patterns of maternal alleles in imprinted regions of the mouse genome at different stages of development. *Mammalian Genome* 13:24–29
- De Rango F, Dato S, Bellizzi D, Rose G, Marzi E, Cavallone L, Franceschi C, Skytthe A, Jeune B, Cournil A (2007) A novel sampling design to explore gene-longevity associations: the ECHA study. *Eur J Hum Genet* 16:236–242
- Dean NL, Loredó-Osti JC, Fujiwara TM, Morgan K, Tan SL, Naumova AK, Ao A (2006) Transmission ratio distortion in the myotonic dystrophy locus in human preimplantation embryos. *Eur J Hum Genet* 14:299–306
- Deng L, Zhang D, Richards E, Tang X, Fang J, Long F, Wang Y (2009) Constructing an initial map of transmission distortion based on high density HapMap SNPs across the human autosomes. *J Genet Genomics* 36:703–709. doi:[10.1016/s1673-8527\(08\)60163-0](https://doi.org/10.1016/s1673-8527(08)60163-0)
- Dudding TE, Attia J, Infante-Rivard C (2004) Addendum to: the association between adverse pregnancy outcomes and maternal factor V Leiden genotype. A meta-analysis. *Thromb Haemost* 92:434
- Dudding TE, Attia J (2004) The association between adverse pregnancy outcomes and maternal factor V Leiden genotype: a meta-analysis. *Thromb Haemost* 91:700–711
- Eaves IA, Bennett ST, Forster P, Ferber KM, Ehrmann D, Wilson AJ, Bhattacharyya S, Ziegler AG, Brinkmann B, Todd JA (1999) Transmission ratio distortion at the INS-IGF2 VNTR. *Nat Genet* 22:324
- Evans D, Morris A, Cardon L, Sham P (2006) A note on the power to detect transmission distortion in parent-child trios via the transmission disequilibrium test. *Behav Genet* 36:947–950
- Eversley CD, Clark T, Xie Y, Steigerwalt J, Bell TA, de Villena FP, Threadgill DW (2010) Genetic mapping and developmental timing of transmission ratio distortion in a mouse interspecific backcross. *BMC Genet* 11:98. doi:[10.1186/1471-2156-11-98](https://doi.org/10.1186/1471-2156-11-98)
- Friedrichs F, Brescianini S, Annesse V, Latiano A, Berger K, Kugathasan S, Broeckel U, Nikolaus S, Daly MJ, Schreiber S, Rioux JD, Stoll M (2006) Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum Genet* 119:305–311
- Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI (2008) Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82:100–112
- Greenwood CM, Morgan K (2000) The impact of transmission-ratio distortion on allele sharing in affected sibling pairs. *Am J Hum Genet* 66:2001–2004
- Haig D, Grafen A (1991) Genetic scrambling as a defence against meiotic drive. *J Theor Biol* 153:531–558

- Hanchard N, Rockett K, Udalova I, Wilson J, Keating B, Koch O, Nijnik A, Diakite M, Herbert M, Kwiatkowski D (2005) An investigation of transmission ratio distortion in the central region of the human MHC. *Genes Immun* 7:51–58
- Hastings IM (1991) Germline selection: population genetic aspects of the sexual/asexual life cycle. *Genetics* 129:1167–1176
- Huang LO, Labbe A, Infante-Rivard C (2013) Transmission ratio distortion: review of concept and implications for genetic association studies. *Hum Genet* 132(3):245–263
- Imboden M, Swan H, Denjoy I, Van Langen IM, Latinen-Forsblom PJ, Napolitano C, Fressart V, Breithardt G, Berthet M, Priori S, Hainque B, Wilde AAM, Schulze-Bahr E, Feingold J, Guicheney P (2006) Female predominance and transmission distortion in the long-QT syndrome. *N Engl J Med* 355:2744–2751
- Infante-Rivard C, Rivard GE, Yotov W, Génin E, Guiguet M, Weinberg C, Gauthier R, Feoli-Fonseca JC (2002) Absence of association of thrombophilia polymorphisms with intrauterine growth retardation. *N Engl J Med* 347:19–24
- Infante-Rivard C, Lévy E, Rivard GE, Guiguet M, Feoli-Fonseca JC (2003) Small babies receive the cardiovascular protective apolipoprotein ε2 allele less frequently than expected. *J Med Gen* 40:626–629
- Infante-Rivard C, Weinberg CR (2005) Parent-of-origin transmission of thrombophilic alleles to intrauterine growth-restricted newborns and transmission-ratio distortion in unaffected newborns. *Am J Epidemiol* 162:891–897. doi:[10.1093/aje/kwi293](https://doi.org/10.1093/aje/kwi293)
- Infante-Rivard C, Rivard GE, Guiguet M, Gauthier R (2005) Thrombophilic polymorphisms and intrauterine growth restriction. *Epidemiology* 16:281–287
- Infante-Rivard C (2010) Genetic association between paraoxonase 1 gene (PON1) single nucleotide polymorphisms and small-for-gestational age outcome in related and unrelated subjects. *Am J Epidemiol* 171:999–1006
- Kryukov GV, Pennacchio LA, Sunyaev SR (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80:727–739
- Lange K (1997) Mathematical and statistical methods for genetic analysis. Springer, New York
- Lemire M, Roslin NM, Laprise C, Hudson TJ, Morgan K (2004) Transmission-ratio distortion and allele sharing in affected sib pairs: a new linkage statistics with reduced bias, with application to chromosome 6q25.3. *Am J Hum Genet* 75(4):571–586.
- LeMaire-Adkins R, Hunt PA (2000) Nonrandom segregation of the mouse univalent X chromosome: evidence of spindle-mediated meiotic drive. *Genetics* 156:775
- Li B, Leal SM (2009) Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genet* 5:e1000481
- Liddell D (1976) Practical tests of 2×2 contingency tables. *J Royal Stat Soc* 25(4):295–304
- Lyon MF (2003) Transmission ratio distortion in mice. *Annu Rev Genet* 37:393–408
- Maher B (2008) Personal genomes: the case of the missing heritability. *Nature* 456:18–21. doi:[10.1038/456018a](https://doi.org/10.1038/456018a)
- Meyer WK, Arbeithuber B, Ober C, Ebner T, Tiemann-Boege I, Hudson RR, Przeworski M (2012) Evaluating the evidence for transmission distortion in human pedigrees. *Genetics* 191:215–232. doi:[10.1534/genetics.112.139576](https://doi.org/10.1534/genetics.112.139576)
- Naumova A, Olien L, Bird L, Slamka C, Fonseca M, Verner A, Wang M, Leppert M, Morgan K, Sapienza C (1995) Transmission ratio distortion of X chromosomes among male offspring of females with skewed X inactivation. *Dev Genet* 17:198–205
- Naumova AK, Greenwood CM, Morgan K (2001) Imprinting and deviation from Mendelian transmission ratios. *Genome* 44:311–320
- Naumova AK, Leppert M, Barker DF, Morgan K, Sapienza C (1998) Parental origin-dependent, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. *Am J Hum Genet* 62:1493–1499
- Novitski E (1951) Non-random disjunction in *Drosophila*. *Genetics* 36:267

- Pardo-Manuel de Villena F, Sapienza C (2001) Nonrandom segregation during meiosis: the unfairness of females. *Mamm Genome* 12:331–339. doi:[10.1007/s003350040003](https://doi.org/10.1007/s003350040003)
- Paterson A, Sun L, Liu XQ (2003) Transmission ratio distortion in families from the Framingham Heart Study. *BMC Genet* 4:S48
- Paterson A, Waggott D, Schillert A, Infante-Rivard C, Bull S, Yoo Y, Pinnaduwa D (2009) Transmission-ratio distortion in the Framingham Heart Study. *BMC Proc* 3(Suppl 7):S51
- Paterson AD, Petronis A (1999) Transmission ratio distortion in females on chromosome 10p11 p15. *Am J Med Genet* 88:657–661
- Polaski A (1998) Dynamic balance of segregation distortion and selection maintains normal allele sizes at the myotonic dystrophy locus. *Math Biosci* 147:93–112
- Riess O, Epplen JT, Amoiridis G, Przuntek H, Schols L (1997) Transmission distortion of the mutant alleles in spinocerebellar ataxia. *Hum Genet* 99:282–284
- Santos PSC, Höhne J, Schlattmann P, König IR, Ziegler A, Uchanska-Ziegler B (2009) Assessment of transmission distortion on chromosome 6p in healthy individuals using tagSNPs. *Eur J Hum Genet* 17:1182–1189
- Sazhenova EA, Lebedev IN (2008) Epimutations of the KCNQ1OT1 imprinting center of chromosome 11 in early human embryo lethality. *Genetika* 44:1609–1616
- Shemer R, Birger Y, Riggs AD, Razin A (1997) Structure of the imprinted mouse *Snrpn* gene and establishment of its parental-specific methylation pattern. *Proc Natl Acad Sci U S A* 94:10267–10272
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506
- Sturtevant A (1936) Preferential segregation in triplo-IV females of *Drosophila melanogaster*. *Genetics* 21:444
- The HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320. doi:[10.1038/nature04226](https://doi.org/10.1038/nature04226)
- Tiwari HK, Barnholtz-Sloan J, Wineinger N et al (2008) Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. *Hum Hered* 66(2):67–86
- Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 62:969–978
- Weinberg CR (1999) Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet* 5(1):229–235
- Westendorp R, van Dunne FM, Kirkwood T, Helmerhorst FM, Huizinga T (2001) Optimizing human fertility and survival. *Nat Med* 7:873
- Wu G, Hao L, Han Z, Gao S, Latham KE, de Villena FPM, Sapienza C (2005) Maternal transmission ratio distortion at the mouse *Om* locus results from meiotic drive at the second meiotic division. *Genetics* 170:327
- Yang L, Andrade MF, Labialle S, Moussette S, Geneau G, Sinnett D, Belisle A, Greenwood CM, Naumova AK (2008) Parental effect of DNA (cytosine-5) methyltransferase 1 on grandparental-origin-dependent transmission ratio distortion in mouse crosses and human families. *Genetics* 178:35–45
- Zhou JY, Hu YQ, Lin S et al (2009) Detection of parent-of-origin effects based on complete and incomplete nuclear families with multiple affected children. *Hum Hered* 67(1):1–12
- Zimmering S (1955) A genetic study of segregation in a translocation heterozygote in *Drosophila*. *Genetics* 40:809
- Zollner S, Wen X, Hanchard NA, Herbert MA, Ober C, Pritchard JK (2004) Evidence for extensive transmission distortion in the human genome. *Am J Hum Genet* 74:62–72

Chapter 13

Epigenome-Wide Association Studies: Potential Insights into Human Disease

Christopher G. Bell

Abstract The burden on human health due to common diseases, such as the metabolic syndrome, cardiovascular and inflammatory disorders is extreme. With increasingly longer lived populations, the morbidity of these chronic conditions leads to vast physical, psychological and economic cost. In order to further understand the pathogenicity of these complex diseases, the intertwined influence of environmental factors and polygenic susceptibility needs to be unravelled. This may at first seem an insurmountably difficult task but progress has been made in recent years.

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
ASE	Allele-specific expression
ASHM	Allele-specific histone modifications
ASM	Allele-specific methylation
BiS-seq	Bisulphite second-generation sequencing
BMI	Body mass index
CGI	CpG island
CTCF	CCCTC-binding factor
DHS	DNase I hypersensitivity sites
DMR	Differentially methylated region
EWAS	Epigenome-wide association study

C.G. Bell (✉)

Medical Genomics, UCL Cancer Institute, University College London, London, UK
e-mail: christopher.bell@ucl.ac.uk

GWAS	Genome-wide association study
HSM	Haplotype-specific methylation
LD	Linkage disequilibrium
LMR	Low methylation region
MeDIP-seq	Methylation-dependent immunoprecipitation second-generation sequencing
MDR	Methylation determining region
MVP	Methylation variable position
RRBS-seq	Reduced representation bisulphite second-generation sequencing
TFBS	Transcription factor binding site
WGAS	Whole genome sequencing association study

13.1 Introduction

The burden on human health due to common diseases, such as the metabolic syndrome, cardiovascular and inflammatory disorders is extreme. With increasingly longer lived populations, the morbidity of these chronic conditions leads to vast physical, psychological and economic cost (Clarke et al. 2010). In order to further understand the pathogenicity of these complex diseases, the intertwined influence of environmental factors and polygenic susceptibility needs to be unravelled. This may at first seem an insurmountably difficult task but progress has been made in recent years.

The advent of Genome-Wide Association Studies (GWAS) (Sladek et al. 2007; The Wellcome Trust Case Control Consortium 2007), built upon population-based linkage disequilibrium (LD) maps (Altshuler et al. 2005) and facilitated by high-throughput array technology, has been highly successful in part of this endeavour. Extensive global scientific work and expenditure has enabled the identification of a large number of common genetic variants, single-nucleotide polymorphism (SNPs), in strong and replicable associations with a wide range of common diseases (see <http://www.genome.gov/gwastudies/>). However, individually these factors only convey a relatively small level of risk and together do not come close to accounting for the estimated heritability of these disorders (Manolio et al. 2009). Additionally the discovery SNPs are rarely found to be code modifiers; approximately 85 % of disease-associated GWAS variants being non-coding (Hindorf et al. 2009). Various potential explanations for this “missing heritable” element have been offered, including methodological factors, such that GWAS do not differentiate between the actual causal variant and another variant in strong LD, the indirect detection of rare and structural variants, and potentially the identification of, or interactions between, stable epigenetic modifications (McCarthy et al. 2008). Also it must be acknowledged that there may be some over-estimation of these heritability measures due to a lack of accounting for genetic interactions (epistasis) (Zuk et al. 2012).

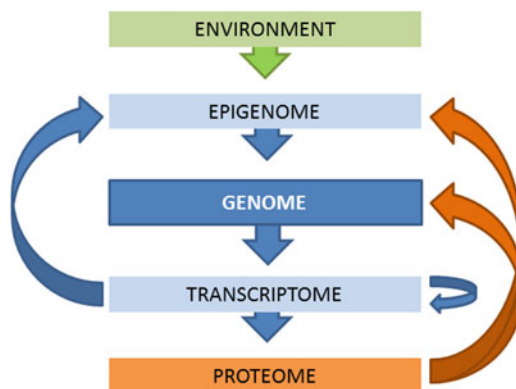


Fig. 13.1 The environment interfaces through the epigenome to influence genome and affect expression. Furthermore non-coding transcripts such as micro-, piwi-interacting-, small interfering- and long non-coding-RNA may influence gene expression by modulation of other transcripts and/or the epigenome (Kaikkonen et al. 2011). Epigenome modifying proteins, such as histone methylators, deacetylators, etc., also influence the structure of the epigenome, thereby influencing expression (Dawson et al. 2012)

The functional interpretation of the robustly replicated GWAS variants discovered to date is the vital next step to fully reap their potential therapeutic and preventative benefits. Even though an individual association only has a small incremental increase in relative risk, solving the biological variation that the allelic sequence difference causes can still identify previously unknown pathophysiological pathways. Recent evidence that many of these GWAS association SNPs are themselves commonly within regulatory DNA has come via genome-wide DNase I hypersensitivity sites (DHSs) experiments. This method detects open chromatin able to be cut by this enzyme and therefore potential functional regions of the genome. This enabled Maurano et al. to identify increasingly stronger associations between association SNPs and DHSs when performed with the most disease appropriate cell types (Maurano et al. 2012), thus indicating the potential functional consequences of these variants when they are located in their correct pathogenic tissues.

Once robust disease-associated genetic variants have been identified via GWAS, or perhaps Whole Genome Association Studies (WGAS) (Jonsson et al. 2012), the medical challenge is to try to decipher how these hereditary components interact with the external environment to control health outcomes. Factors such as stress, diet and toxins may be translated into biological effectors on the genome via epigenetic changes (Feinberg 2008; Bell and Beck 2010) (Fig. 13.1). Therefore these genetic effects do not act in isolation. It is how these external factors may influence these disease susceptibilities, how they may interface with the genome via the epigenome (Bell and Beck 2010) and whether there is tractable ability to measure this effect that may bring further breakthroughs in the understanding of disease pathogenesis.

13.2 The Epigenome

There are greater than 200 distinct cell types within the human body (Alberts et al. 2008) and all these cells possess the same genome, barring somatic mutation. It is the packaging and chemical modifications of the ~2 m length of DNA within each (Strachan and Read 1999) that influences gene expression and therefore enables tissue-specific activity. The components of this overlying mechanism are termed the “epigenome”. This includes modifications of the DNA molecule itself, as well as alternates to the histone proteins that the helix winds around, including amino acid variants to the protein structure and post-translational modifications of the appending tails of these molecules. Also certain species of non-coding RNA, such as micro-, piwi-interacting-, small interfering- and long non-coding-RNA may be considered as part of the epigenome, due to their influence on gene expression (Kaikkonen et al. 2011).

A steadily expanding list of epigenetic modifications are now known to exist including currently four DNA modifications: DNA methylation (5mC), hydroxymethylation (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Branco et al. 2012); 16 classes of histone tail modification (Dawson and Kouzarides 2012; Tan et al. 2011) leading to more than 200 alterations; and histone variants such as H2A.Z and H3.3 (Conerly et al. 2010; Jiang and Pugh 2009) (see Table 13.1). The 5-methyl modification of cytosine also is found to occur in RNA (Squires et al. 2012) and recently a N6-methyladenosine (m⁶A) modification of RNA was characterised genome-wide (Dominissini et al. 2012; Meyer et al. 2012), of which the *FTO* gene functions as its demethylase (Jia et al. 2011).

The strict definition of a “true” epigenetic mechanism requires the modification to be able to be propagated through mitotic replication and therefore to be mitotically heritable. The robust mechanism for this inheritance is well understood for DNA methylation, though less certain for other modifications (Bird 2002). Three DNA methyltransferase enzymes classically act as maintenance (DNMT1) and de novo methylators (DNMT3A & B), respectively. The maintenance enzyme, DNMT1, acts by recognising hemi-methylated DNA (Margueron and Reinberg 2010), although this orthodox understanding may in fact be an oversimplification with all three potentially now thought to play a role in this process (Jones and Liang 2009).

In order to investigate the epigenome there needs to be clarity as to how this layer of information differs from the underlying genome. This can be broken down into four principle facets. Firstly, it is by function tissue-specific (Fernandez et al. 2012; Bock et al. 2012). Secondly, it is not static but changeable over time, due to developmental programming alterations (Ferguson-Smith 2011), random drift (Feil and Fraga 2011), age-related modifications within specific loci (Teschendorff et al. 2010; Rakyan et al. 2010) or environmental influence (Feinberg 2007). Thirdly, it can vary between homologous chromosomes, in either a parent-of-origin-specific manner that occurs within imprinted loci, required for normal development such as the Prader-Willi locus (Ferguson-Smith 2011), or other non-imprinted allele-specific means. This will manifest as allele-specific methylation

Table 13.1 Epigenetic modifications and marks (Dawson and Kouzarides 2012; Conerly et al. 2010; Jiang and Pugh 2009; Squires et al. 2012; Dominissini et al. 2012; Meyer et al. 2012)

DNA modifications	5-Methylcytosine	5mC
	5-Hydroxymethylcytosine	5hmC
	5-Formylcytosine	5fC
	5-Carboxylation	5caC
Histone tail modifications	Acetylation	K-ac
	Methylation (lysine)	K-me1, K-me2, K-me3
	Methylation (arginine)	R-me1, R-me2s, R-me2a
	Phosphorylation (serine and threonine)	S-ph, T-ph
	Phosphorylation (tyrosine)	Y-ph
	Ubiquitylation	K-ub
	Sumoylation	K-su
	ADP ribosylation	E-ar
	Deimination	R
	Proline isomerisation	P-cis
	Crotonylation	K-cr
	Propionylation	K-pr
	Butyrylation	K-bu
	Formylation	K-fo
	Hydroxylation	Y-oh
	O-GlcNAcylation (serine and threonine)	S-GlcNAc; T-GlcNAc
Histone variants	H2A.Z	H2A.Z
	H3.3	H3.3
RNA modifications	5-Methylcytosine	5mC
	N6-methyladenosine	m6A

(ASM) or allele-specific histone modifications (ASHMs). Finally the epigenome can be influenced by genetic polymorphism (Kerkel et al. 2008; Shoemaker et al. 2010). This can drive differences between individuals or may also contribute to the non-imprinting related allele-specific variation between homologous chromosomes.

The tissue-specific epigenome of a particular cell type of an organ, for example the liver or hepatocyte epigenome, will be more different from a skin cell within an individual than two hepatocyte epigenomes between two individuals, i.e. intra-individual inter-tissue variation exceeds inter-individual variation in any given tissue (Davies et al. 2012). In fact the tissue-specific epigenome even between two related species, such as human and chimpanzee, will be more similar than when compared with a different tissue or cell type within the same individual (Pai et al. 2011). However these tissue-specific templates will not be precisely identical as genetic variation and other potential environmental influences lead to subtle modifications between people (Brena et al. 2006). It is this fraction of variability upon the underlying rigid tissue-specific design that we are most interested in for disease studies, and hope to dissect within this the proportion that can be attributed to genetic variability, and that which is arising from environmental influences.

Additionally if we wish to perform studies on the epigenome, we need to be well aware that there is high potential for random artefacts to be found if cell lines are used as the analyte, due to the stochastic changes that occur due to artificial environment and procedures experienced by these cells (Brennan et al. 2009; Aberg et al. 2012). Similarly we need to recognise that whole blood derived DNA is from a mixture of blood cell types, so the resultant epigenetic findings will be a proportional representative epigenome of all of the constitutive blood type cells present.

13.3 DNA Methylome

DNA methylation, the addition of a methyl group to the 5-carbon of cytosine, is currently the most well-studied epigenetic modification. Within the human genome this occurs almost exclusively in differentiated cells in the context of a CpG dinucleotide: cytosine followed by guanine from 5' to 3' in the DNA strand via the phosphodiester bond. Normal embryonic development requires the ability to methylate these cytosines, as this mechanism is involved in genomic imprinting, X chromosome inactivation, and repression of transposable elements differentiation (Robertson 2005). There are approximately 28 million CpG dinucleotides within the human genome and around 70–80 % of these CpGs are methylated in mammals (Lister et al. 2009). Those CpG dinucleotides that remain unmethylated predominantly occur in clusters, termed “CpG islands” (CGI), which sparsely punctuate the genome. These islands number 27,718 or 22,374 via UCSC or Ensembl databases, respectively, with discrepancy in number due to differing repeat masking and size requirements in the respective prediction algorithms (Illingworth and Bird 2009). CGI account for ~7 % of all genomic CpGs. These clusters when unmethylated recruit CpG binding proteins, chromatin modifying enzymes, such as Cfp1 and KDM2A, leading to the modification of histone tails (Blackledge et al. 2010; Thomson et al. 2010) and permissive chromatin formation. Therefore CpG islands can be described as a “platform” on which gene transcription is able to occur (Blackledge and Klose 2011). Approximately half of CGI are found at the promoter region of genes, with the remainder evenly split between intergenic and intragenic locations (Illingworth et al. 2010). Of these non-promoter CGI, those located elsewhere within genes may potentially act as alternate isoform promoters (Illingworth et al. 2008; Maunakea et al. 2010) and some of those between genes may be involved in currently unidentified non-coding transcripts.

The islands stand out as the vast majority of the genome is depleted of CpGs dinucleotides. There are in fact ~120 million of the exact reverse dinucleotide GpC. This reduction is due to the hypermutability of methylated cytosines (Duncan and Miller 1980; Roach et al. 2010) which are easily deaminated to uracil and subsequently converted to thymine. Therefore where cytosine methylation occurs within the germ-line cells of sperm or egg, there will be a high chance of mutational loss of these CpGs over evolutionary time. These dinucleotides will be converted to TpG

or CpA depending on the strand of the cytosine deaminated. This is well demonstrated by the fact that 18.2 % of human pathological lesions were documented by Cooper et al. to be C to T or G to A transitions located within a CpG context (Cooper et al. 2010); transitions at CpGs occur at $>\sim 13$ times non-CpG transitions, and so play a significant role in SNP occurrence (Li et al. 2009). This CpG-SNP variability has been shown to have considerable influence on the surrounding methylome and therefore strongly contributes to ASM variability (Shoemaker et al. 2010).

Dramatic DNA methylome changes occur in cancer genomes, such as a global hypomethylation with concurrent de-repression of pathogenic repetitive elements and locus hypermethylation of tumour suppressor genes (Esteller 2007). However with the increased resolution in methylome analysis now possible, initially with array technology but increasingly with second-generation sequencing, more subtle signatures have been identified. This includes findings that more differentially methylated regions (DMRs) are likely to be seen in regions surrounding islands approximately 2 kb up- and downstream, termed “CpG Island shores” than within the CpG islands themselves (Irizarry et al. 2009). These regions were found to contain the most significant tissue-, cancer- and reprogramming-specific changes (Doi et al. 2009) and in malignant tissue were shown to be contributed to by the loss of the strictly delineated methylation change, or boundary, at island borders (Hansen et al. 2011). High resolution analysis in human and chimpanzee sperm has identified islands to have wider regions of hypomethylation, stretching further into their shore regions, within these germ-line cells (Molaro et al. 2011). These larger hypomethylated islands and shores regions are shown to recede to different degrees during differentiation and this tidal nature of methylation around islands can also be seen clearly in the process of haematopoietic delineation (Hodges et al. 2011). The first human bisulphite sequencing methylomes, published by Lister et al., performed in embryonic stem cells and fibroblasts also identified intermediate methylation regions, defined as partially methylated domains (PMDs), as being disproportionately lost in the differentiation process (Lister et al. 2009).

DNA methylation modification has location-specific functional effects depending upon where within the DNA code it resides, i.e., repressive within CpG Islands and CpG Island Shores, activating within Gene Bodies (Hellman and Chess 2007), or potentially repressive within transcription factor binding sites (TFBSs). Whilst clusters of CpG dinucleotides may act in concert, the CpG dinucleotide can also be thought of as a genomic signalling molecule in its own right (Bird 2011). The presence or absence of an individual CpG within TFBSs facilitating methylation may influence the variability of binding, as seen in recent CTCF data (Wang et al. 2012). In this ENCODE study 41 % of the variability of CTCF occupancy was associated with differential DNA methylation. Sites that showed more prevalent methylation-associated variability were enriched for the genetic occurrence of a CpG dinucleotide at two loci within the CTCF motif (Fig. 13.2). Their presence allowed methylation variability to influence occupancy by ~ 2 times compared to motifs lacking these CpGs and were termed CTCF

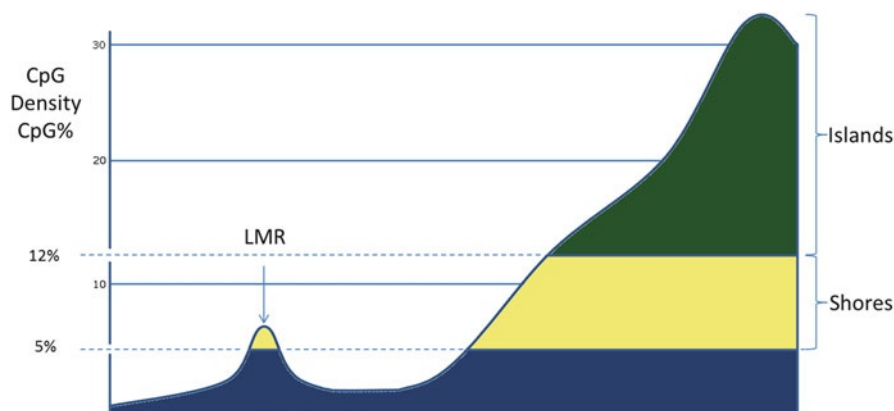


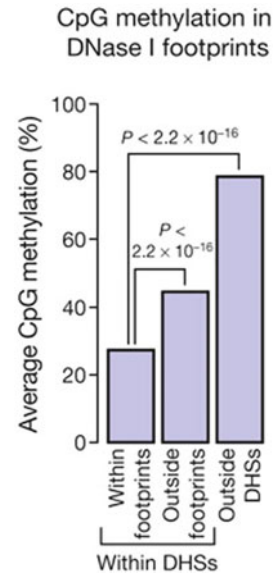
Fig. 13.3 CpG density across the genome. CpG islands possess density of >12 % up to over 30 % CpG density (i.e., 100 % CpG = 50 CpGs in 100 bp). CpG island shores (Irizarry et al. 2009) form an intermediate level of density beside islands before dropping back to genome baseline levels of around ~2 %. Low methylation regions (LMRs) (Stadler et al. 2011) are predominately promoter distal regions that rise above this genomic ocean baseline

interaction. Stadler et al. in a mouse methylome experiment identified distal regulatory regions with enhancer chromatin signatures that had above genome baseline CpG densities of ~2.5–5 %, included about 4 % of all mouse CpGs, and possessed a low methylation, averaging 30 % (Stadler et al. 2011). These loci were termed low methylation regions (LMRs) (Fig. 13.3). They observed the binding of cell type-specific transcription factors to these regions led to loss of methylation. Passive deposition within cell-specific TFBSs after their vacation was also supported by Thurman et al. within DNAase I hypersensitivity sites (DHS) (Thurman et al. 2012).

Although even if methylome variability within these binding sites is passive, it still remains informative of stable DNA–protein interactions. Supporting this is recent data on TFBS “footprints” that were identified within the larger DHS sites, as their binding protected this small motif region from DNase I cleavage (Neph et al. 2012a). A connection between regulatory factor occupancy and base-resolution DNA methylation status was shown, with decreased methylation within DHS sites, but moreover within these regions, it is significantly less within the motif “footprints” than outside them (Neph et al. 2012a) (Fig. 13.4). Measures of DNA methylation additionally have the advantage of a robust replication system and slower turnover compared to the potential temporality of expression. Equally we need to consider how a completely passive role would reconcile with the long-standing evidence of DNA methylation repressing transcription in *in vivo* transcriptional silencing (Siegfried et al. 1999).

Genetic influences on the methylome can be split into those caused directly by the CpG dinucleotide and other polymorphic *cis*- or *trans*-effectors that influence the methylation machinery. Density change within CGIs affects their dynamic ability to methylate, with the vast majority of high-density CGI unmethylated

Fig. 13.4 Figure from Neph et al. (2012a). Reprinted by permission from Macmillan Publishers Ltd: Nature 489:7414 © 2012. Significant methylation variability from outside DNase I hypersensitivity sites to within, and within between transcription factor footprints of binding and outside this footprint. (data from IMR90 cells, $P < 2.2 \times 10^{-16}$, Mann–Whitney U -test)



irrespective of their transcriptional state, and low-density CGI the preferable template for tissue-specific methylation (Weber et al. 2007; Eckhardt et al. 2006). In addition within CGI, proposed TFBSs influence the likelihood of methylation, which are termed methylation determining regions (MDRs) (Lienert et al. 2011).

Allelic SNP variation may be *cis*-effectors within regulatory motifs enabling methylation (Kerkel et al. 2008; Tycko 2010; Schalkwyk et al. 2010). This ASM may contribute to and highlight potential allele-specific expression (ASE) patterns (Bell and Beck 2009). Quantitative trait loci (QTL) for DNA CpG methylation have been identified in the brain tissue (Zhang et al. 2010) with *cis* peaks only 45 bp from the focal CpG site and were more likely to occur for non-CGI CpGs (Gibbs et al. 2010). Fang et al. identified in humans that regions of ASM found in multiple cell types were frequently the promoters of non-coding RNA (Fang et al. 2012). Potential global *trans*-effectors on methylation have also been identified, i.e., rs10876043 in *DIP2B* (Bell et al. 2011).

Environmental factors including nutrition, behaviour, stress, toxins, as well as stochastic factors, have also been found to affect the epigenome (Faulk and Dolinoy 2011). Therefore because of both DNA methylation's replication stability, but also its potential plasticity, it is proposed as a biomarker of quantitative lifetime environmental exposure or accrued pathogenic alteration (Feinberg 2007; Bock 2012).

13.4 Variation in the Methylome

Differences identified in methylomes may depend upon methodology implemented, but terminology needs to be accurately understood in order to clearly appreciate the significance of findings. Base-resolution analysis will enable the identification of

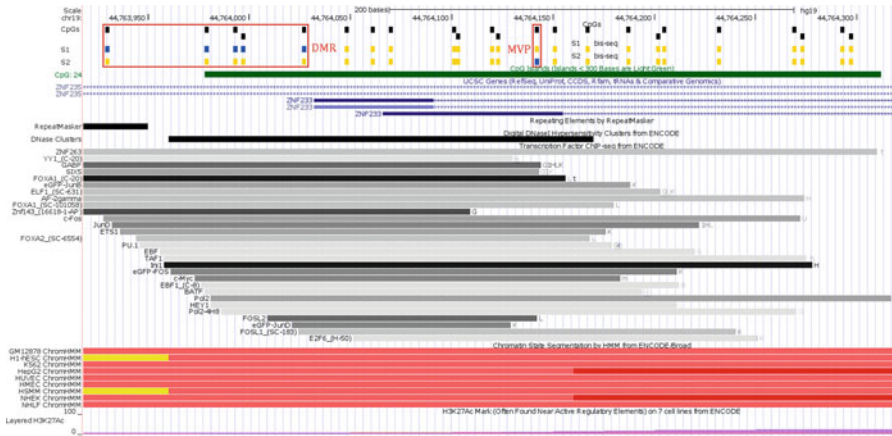


Fig. 13.5 UCSC browser with mock data showing methylation variability. Yellow CpG $\leq 20\%$ and blue CpG $\geq 80\%$ methylated. CpG methylation data can be compared (here between Subject 1 (S1) and Subject 2 (S2)) and variation can be isolated to a single CpG or methylation variable position (MVP) or be a regional difference spread over a number of CpGs—a differentiated methylated region (DMR). These MVPs or DMRs can then be assessed to see if they reside within an annotated genomic feature, i.e., CpG islands, CGI shore, transcription factor binding site, chromatin defined state (ChromHMM (Ernst and Kellis 2012) see Table 13.2), as shown here within an active promoter region, as defined by ChromHMM in multiple cell types, with potential influence on a number of transcription factor binding sites

methylation variable positions (MVPs) which represent variation at individual cytosine. DMRs indicate a regional difference, this may be from comparison with an affinity-based method such as MeDIP, a number of contiguous CpGs, regional-sized window, or within an annotated genomic feature (such as CGI, CGI Shore, TFBS or Enhancer) (Fig. 13.5).

Differences between homologous alleles may be defined, by methylation variation at an exact individual cytosine, as ASM. Larger regions may be termed haplotype-specific methylation (HSM) (Bell 2011), which may be driven genetically by the co-ordinated phase of CpG-SNPs or include *cis*-genetic as well as pure epigenetic differences and form a “hepitype” (Murrell et al. 2005).

13.5 EWAS: Epigenome-Wide Association Studies

The advent of high-throughput DNA methylation arrays, such as the Illumina Infinium 450 k which interrogates DNA methylation at over 480,000 individual CpG sites (Bibikova et al. 2011; Dedeurwaerder et al. 2011), enables large-scale epigenomic studies to be performed. This technology uses bisulphite conversion to render the cytosine at CpG loci into a pseudo-SNP depending on its methylation status, thereafter comparable Illumina SNP typing technology is implemented.

The undoubted success of GWAS has unfortunately led many to hopefully assume that the same study design methods would be equally successful in the Epigenome-Wide Association Study (EWAS) context, were associations are tested for between a phenotype and genome-wide epigenetic marks. However the principle differences between the genome and epigenome, as previously discussed, all contribute to make case-control population-based studies in whole blood-derived DNA a less optimal design than in GWAS.

The first EWAS studies performed in common diseases, including investigations in systemic lupus erythematosus (Javierre et al. 2010), type 1 diabetes nephropathy (Bell et al. 2010a), autism spectrum disorders (Nguyen et al. 2010), major psychosis (Mill et al. 2008) and body mass index (BMI) (Feinberg et al. 2010) were able to identify methylation changes, although uniformly very small. However further external replication will definitely be required and we need to be very conscious of the “winner’s curse” of many genetic association studies, prior to the advent of rigorous GWAS methodology, that were subsequently unable to be replicated (Bell et al. 2007). At present we do not have a definitive and mature methodology for EWAS and whether the current technological advances will be sufficient, or if further will be required for a significant breakthrough to be made, cannot yet be stated (Heijmans and Mill 2012).

The largest effects that have been identified to date have been due to the extreme environmental influence of tobacco smoking on the methylome. An early study using the previous Illumina 27k array performed a smoking EWAS using peripheral blood and identified a ~12 % difference in methylation in a CpG, cg03636183, located in *F2RL3* (coagulation factor II (thrombin) receptor-like 3). This was able to be detected in small number of 65 heavy smokers (interquartile methylation range 78–88 %) and 56 non-smokers (94–96 %), with replication in an additional 316 independent samples and therefore was small by GWAS scale comparisons (Breitling et al. 2011). Obvious significant potential confounders exist; including genetic heterogeneity and mixed cell type-derived DNA. In a later study now with the 450 k array in 1,062 new-born cord blood samples from the Norwegian Mother and Child Cohort Study (MoBa) an association with plasma cotinine levels, a biomarker of tobacco intake, by mothers during pregnancy and median methylation levels at 26 CpG located in ten genes was found (Joubert et al. 2012). The higher resolution analysis enabled more convincing clustered CpG evidence to be identified, with three distinct loci with at least four CpGs, having the same directional change. Replication in offspring cord blood samples was surprisingly possible in only 18 smoking exposed versus 18 non-smoking pregnancies, which all had the small directional change in methylation and 21 of the 26 CpGs were significant to the level of $P < 0.05$ and 5 yet still to post-Bonferroni correction level. The aryl hydrocarbon (AhR) signalling pathway known to be involved in the detoxification of tobacco smoke was highlighted by two genes, *AHRR* and *CYP1A1*, by eight implicated CpGs in this *in utero* exposure. Furthermore this study did individually support the previous cg03636183 *F2RL3* finding at a $P < 0.05$ level. Also to exclude major cell type variation contributing to this result, the replication samples were separated into the major cell types of mononuclear and polymorphonuclear

cells, and median methylation variation between cell type was an order of magnitude less than seen for maternal smoking. One of the CpGs, cg14817490, in *AHRR* was also identified to have the same directional significant lower methylation in an EWAS in adult smokers (Monick et al. 2012). Additionally that finding was replicated in both B lymphoblastoid cells and also alveolar macrophage cells, therefore the initial finding had not been confounded by cell type mixtures. Moreover this *AHRR* finding (cg05575921) has now also been replicated in youths with a stepwise effect even from low levels of smoking (Philibert et al. 2012). These initial successes in smoking EWAS indicate potentially that if the effect is strong enough confounding factors may be overcome, but this may also reflect what an extreme environmental influence this carcinogenic mixture is.

Current practical reasons limit high-throughput epigenomic analysis to the DNA methylation mark (Heijmans and Mill 2012) although this may rapidly change. Recognisable study design modifications to improve EWAS power include reducing the genetic influence by using monozygotic twins that are discordant for the disease studied (Bell and Spector 2012), if these are at all available, or longitudinal studies over time within the same individuals, which again are rare resources (Rakyan et al. 2011) (see also Chap. 14). Comparison between monozygotic and dizygotic twins can be powerful in delineating genetic and potential environmental causes (Bell and Spector 2012). Further difficulty comes in that due to tissue specificity, it is obviously highly desirable to use the pathogenic tissue involved in the disease process. Although a recent study from Davies et al. showed the possibility that some of the variation between individuals was reflected across both brain and blood, indicating that the use of surrogate readily accessible peripheral tissues may have limited utility at a subset of loci (Davies et al. 2012). Variation identified across all soma tissues could stem from genetic influence, extremely early induced developmental epigenetic modulation subsequently maintained through all or most cells, such as putative human metastable epialleles (Waterland et al. 2010), or environmental influences that lead to signatures across multiple cell types whilst only leading to pathogenicity in a specific tissue. With the majority of epigenome being established during embryogenesis and early foetal development (Squires et al. 2012), this makes this a critical window for potential environmental effects (Gluckman et al. 2008).

The non-static nature of the epigenome also needs to be considered for potential confounding effects on age. As mentioned specific loci have been implicated in ageing, so correction within these identified regions in the promoters of polycomb group proteins target genes and bivalent domain loci will be required (Teschendorff et al. 2010; Rakyan et al. 2010). Random drift over an adult lifetime (Heyn et al. 2012) and developmental changes that also continue into childhood need to be accounted for. For example changes in monocytes in the first 5 years of life, with possible link to immune maturation, having been identified (Martino et al. 2011).

Another potential confounder is the use of heterogeneous DNA from whole blood, although the simple integration with blood cell count data may enable correction for dramatic change due to effects such as infection (Heijmans and Mill 2012). In fact extrapolation of the constitutional cellular fractions from only

methyome data has been recently described (Houseman et al. 2012) and will only become more accurate with more detailed data. Finally, if robust changes are identified, longitudinal disentanglement is required to determine if they are in fact causative in the pathology, or an effect of the disease, or even the treatment or medication regime given (Heijmans et al. 2009). Validation by an alternate method remains desirable although currently bisulphite conversion, as implemented within the Illumina arrays protocol is the gold standard. This will quickly change if the possibilities of third-generation sequencers to directly read DNA modifications come to fruition (Flusberg et al. 2010; Clarke et al. 2009).

Epigenetics may play a causative role in complex diseases, but it will not be contributing to “missing heritability” (Slatkin 2009) unless it has a facultative or obligatory relationship with genetics (Richards 2006). Therefore amalgamation of these analyses may enable further insight into these common diseases. A strategy is to utilise the considerable strength of GWAS and look for subtle epigenomic variability within the regions robustly confirmed to have some contribution to the disease process (Birney 2011; Bell 2011). Identified epigenetic factors may be ratcheting-up or -down the underlying genetic influence. If these methylation changes, for instance, are an environmentally driven modification, this would be a mechanism for genetic and lifestyle factors to combine their influence. By integrating epigenomic information, including DNA methylation and chromatin data, within regions of genetic susceptibility, this may enable insight into the functional mechanism, modes of inheritance and potential environmental modulation (Birney 2011).

13.6 Integration of Epigenomic and Genomic Data

Increased or attenuated effects of common genetic predisposition have been identified due to environmental stimuli. These modulators are likely to function through epigenomic pathways. Examples include the *FTO* locus, the strongest common susceptibility variant associated with BMI (Dina et al. 2007; Frayling et al. 2007). Like most GWAS association hits, the identified *FTO* association SNPs lie within a non-coding region, in this case in the first intron of the gene. In a large study of over 200,000 adults from Kilpelainen et al. the effects of the susceptibility allele were lessened by physical activity, reducing the odds of obesity from 1.3 per allele to 1.22 (Kilpelainen et al. 2011). Additionally in a recent study from Qi et al. the genetic predisposition for BMI was calculated from 32 known associated SNPs and in over 30,000 individuals the genetic association was found to be more pronounced in those with higher intake of sugar-sweetened beverages than those with a lower level (Qi et al. 2012).

One method to integrate methyome and genomic data is termed HSM analysis, whereby DNA methylation is measured within the LD block of a disease association SNP (Bell 2011) (Fig. 13.6). The methylation levels are then analysed not by case versus control, but by risk haplotype status and a linear relationship

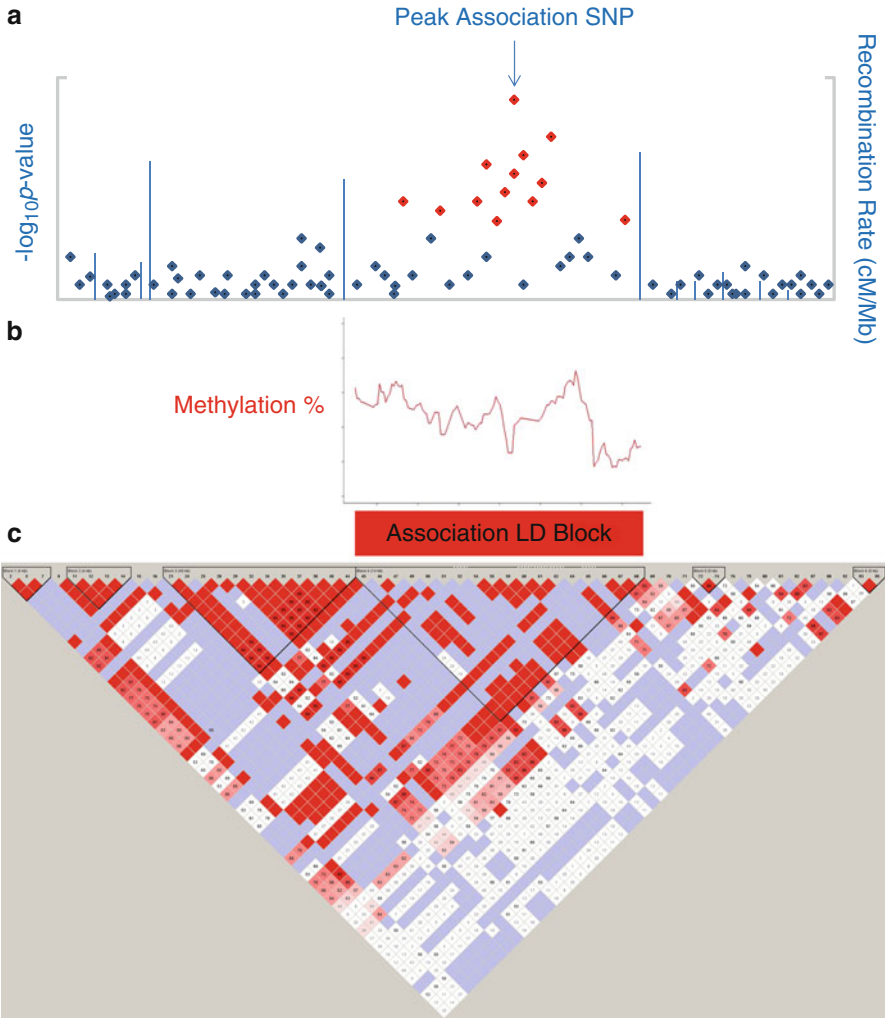


Fig. 13.6 Haplotype-specific methylation analysis (from Bell (2011), Adapted from Personalized Medicine, May 2011, Vol. 8, No. 3, Pages 243–251 with permission of Future Medicine Ltd). DNA methylation is calculated across the linkage disequilibrium (LD) block of disease-associated SNP. (a) Illustration of a GWAS peak locus with the X -axis representing $-\log_{10} P$ -values for SNPs (points: red above significance threshold; blue non-significant) and local recombination rate as blue lines (cM/Mb) with peaks demarking the LD blocks represented (c) in Haploview (Barrett et al. 2005). (b) Figure of the average DNA methylation identified across the association LD block locus

between the three genotypic groupings (homozygote risk, heterozygote and homozygote non-risk) is looked for. This can be investigated for across the entire LD block for potentially strong and co-ordinated effects or a sliding window analysis with variable-sized windows can be performed to isolate potentially

stronger effects and functional regions. The power of this integrative approach is that it uses the strength of the large-scale GWAS studies performed in 1,000's of samples, and then looks for subtlety between these alleles according to their epigenetic state. This analysis was performed in all then known association regions for Type 2 Diabetes in a study in 2010 (Bell et al. 2010b). A significant relationship between methylation state and the *FTO* locus was identified, whereby the obesity susceptibility haplotype possessed significantly higher levels of methylation. This was localised within a 7.7 kb region and the epigenomic architecture here revealed it to possess a chromatin histone H3K4me1 enhancer signature in tissues including skeletal muscle. The most extreme difference of ~10 % methylation between homozygote risk and non-risk carriers was then isolated in a narrow peak of 0.9 kb. This was found to be strongly contributed to by the co-ordinated phase of CpG-creating SNPs across the risk haplotype within this region, tagged by the CpG-SNP rs7202116. The utility of this approach can be shown in that statistically significant results were identified, across a 47 kb LD block, using just the methylation data of 60 individuals. Thousands of individuals would have been required to identify this as a disease-related case versus control DMR without input of the genotype information.

Subsequently Yang et al. performed a meta-analysis of approximately 170,000 samples to try to identify SNPs associated with a trait's phenotypic variability (Yang et al. 2012). This identified that the methylatable CpG-SNP rs7202116, within the obesity-related *FTO* LD block locus, was strongly associated with BMI phenotypic variance (Yang et al. 2012). This was the "G" allele of this SNP that tagged the higher region of genetically driven methylation difference in the study above and itself created a CpG site. DNA methylation was postulated to be the possible functional cause (Yang et al. 2012). Therefore a potential hypothesis to connect these two findings would be that the increased variability was being facilitated by the presence of a CpG, or number of CpGs, within a particular functional region such as a TFBS. As has been proposed in the CTCF binding data from Wang et al. (2012), methylation within this region is coupled with binding and therefore downstream functional consequences. Consequently the CpG sites facilitate higher variability by their ability to subtly modify methylation, perhaps in a tissue- or developmental-specific timeframe, whilst the non-CpG possessing allele only has a more rigid set point response and consequently reduced variance.

Therefore the genetic variability due to CpG gain or loss may be an important facilitatory factor for environmental influence and this role may be predicted from phased haplotype data and methylome analysis in any tissue. The strong influence of genetic polymorphism from CpG-SNPs on the methylome has been identified, as mentioned (Shoemaker et al. 2010), though the functional relevance will be wide-ranging depending upon genomic location. The tissue specificity of the epigenome may then be acting above or facilitated by this.

Further integration with chromatin data also aids the functional dissection of these genetic regions. The co-ordination of histone modification code data, as well as binding sites of the insulator protein CTCF, has enabled broad scale chromatin

Table 13.2 Annotation of the Genome from chromatin signatures (ChromHMM (Ernst and Kellis 2012; Ernst et al. 2011))

State	CTCF	H3K27me3	H3K36me3	H4K20me1	H3K4me1	H3K4me2	H3K4me3	H3K27ac	H3K9ac
Active Promoter									
Weak Promoter									
Inactive/Poised Promoter									
Strong Enhancer									
Strong Enhancer									
Weak/Poised Enhancer									
Weak/Poised Enhancer									
Insulator									
Transcriptional Transition									
Transcriptional Elongation									
Weak Transcribed									
Polycomb Repressed									
Heterochromatin - Low signal									
Repetitive/CNV									
Repetitive/CNV									

Freq %	
<10 %	
11-20	
21-40	
41-60	
61-80	
81-90	
>90	

states to be predicted (ChromHMM) (Ernst and Kellis 2012; Ernst et al. 2011) (see Table 13.2). This may be used to delineate functional relevance of methylome changes within these chromatin delineated regions, such as promoters, transcribed gene bodies or enhancers, e.g., potential influence on eRNA (Zhou et al. 2011a; Kim et al. 2010). The integration of GWAS haplotypes with chromatin state has also begun to be explored within nine cell types (Ernst et al. 2011) and available in the HaploReg database (Ward and Kellis 2012). Subtle methylation variability influencing TFBSs may be highly informative, if powerful-enough studies can be performed to detect this. Encode data of TFBSs in many cell types are now available (Neph et al. 2012a, b; Spitz and Furlong 2012). As well international consortium projects, IHEC and Roadmap, have increasing sets of publically available epigenome datasets available now and to come in the near future (Zhou et al. 2011b).

13.7 Methylome Analysis

DNA methylome biological analysis can be split into three major techniques: affinity enrichment by use of antibodies, such as methylated DNA immunoprecipitation (MeDIP); enzymatic by the use of methylation-sensitive restriction enzymes; or chemical conversion, usually the bisulphite (BiS) reaction (Schones and Zhao 2008). All three can be used either to investigate an individual genetic locus, be coupled with arrays for high-throughput analysis, or with second-generation sequencing for genome-wide high resolution methylome analysis.

The most common examples for these sequencing based combinations include MeDIP-seq, RRBS-seq (Reduced Representation BiSulphite sequencing) and BiS-seq (BiSulphite sequencing). Dependent upon what biological question is being asked and obvious cost implications, the various methylome analyses have different advantages and disadvantages (Robinson et al. 2010). MeDIP obtains good coverage of intermediate methylation regions, such as CGI shores, per read mapped though without single CpG resolution. RRBS enables single CpG resolution but data is largely only available in the CpG dense CGI regions. The current gold standard is whole genome shotgun BiS-seq, with the first human methylomes published using this technique in 2009 (Lister et al. 2009). This study identified a considerable level of non-CpG cytosine methylation in embryonic cells that was lost with differentiation. The cost of and the read requirements of BiS-seq are still extreme, requiring $>10^9$ reads. In published BiS-seq studies to date in human, average alignable genomic coverage of $\sim 28\times$ (Lister et al. 2009), $24.7\times$ (Li et al. 2010), $34.3\times$ (Zeng et al. 2012) and $16\times$ in sperm (Molaro et al. 2011) was reached. Another possible consideration to reduce cost is whether a pooled analysis could be appropriate.

The computational analysis of genome-wide methylome data obviously has to be highly tailored by the analysis technique used to generate the data, with examples including bespoke pipelines for MeDIP-seq data, MeDUSA (Wilson et al. 2012) which incorporates MEDIPS (Chavez et al. 2010), and other detailed biological and bioinformatics BiS-seq analysis protocols (Johnson et al. 2012). The numerous computational methods have been recently reviewed excellently by Bock (2012). Many BiS-seq protocols to date have not adequately dealt with the issue of SNPs, particularly the significant issue of C > T SNP/unmethylated cytosine ambiguity (Liu et al. 2012). Although a BiS-seq SNP caller has recently been published (Liu et al. 2012) and identified correctly 96 % of SNPs with $\sim 30\times$ coverage.

Whilst the identification of an isolated CpG variation, or MVP, in a genome-wide data set cannot be excluded as non-critical, it has more risk of false positivity especially with the large number of CpGs probed, as well as being functionally more difficult to delineate. Therefore regional DMR methods using this individual cytosine data from array or single resolution sequencing have been developed, such as the “bump hunting” methodology (Jaffe et al. 2012) to look for statistically outlying regions of methylation and the CpG_MPs analysis tool (Su et al. 2012) which uses a hotspot extension algorithm to identify unmethylated and methylated regions.

13.8 Evolutionary Aspects in the Analysis of the Epigenome

Human uniqueness in both our phenotype but also disease susceptibility is the result of interaction between the genome, environment, behaviour and culture (Varki et al. 2008). Acquired epigenetic changes are highly significant in the evolution of the malignant genome (Esteller 2007), and these forces can be similar over the different timeframes of cell, population and species evolution (De and Babu 2010). Comparative genomic analysis between species has been successful in identifying potential functionally significant regions due to long-standing conservation (Lindblad-Toh et al. 2011). However, comparative epigenomic analysis adds additional layers of complexity, as species-specific genetic, epigenetic and environmental influences may be driving any differences found. Therefore the epigenomic variation identified, within the same tissue between highly similar species, may be a mechanism to dissect out these factors, and furthermore delineate the loci in which this variability occurs. In regions of strong sequence similarity there is a higher probability of this change being environmentally driven, although potential long range *cis*- or *trans*-acting genetic differences cannot be excluded.

The comparative epigenomic approach to date has included vertebrate ChIP-seq of TFBSs (Schmidt et al. 2010) and primate methylome analyses (Molaro et al. 2011; Martin et al. 2011; Pai et al. 2011). Comparative BiS-seq of frontal cortex between human and chimpanzee showed age- and sex-related changes, but also that differentially methylated genes were enriched for neurological, psychological and cancer-related disorders (Zeng et al. 2012).

With the identification of human-specific DMRs, a subset will possess significant environmentally driven differences. More subtle potential *intra*-human deviation in these known variable regions may identify environmental effects that could be tested for association with disease, such as switched-up or -down inflammatory or developmental pathways. Interestingly, human species-specific DMRs in a comparative tri-primate (human, chimpanzee and macaque) analysis, by MeDIP-seq in peripheral blood, were found to be more prevalent in CGI shores than the islands themselves (Wilson et al. in preparation), in a similar fashion to cancer and tissue DMRs (Irizarry et al. 2009).

Genetic change in CpG presence may influence CpG island density, or lead to the erosion of or accretion within shore regions, thus influencing the location of these highlighted regions of potential variability. As discussed the necessity of this dinucleotide template to facilitate methylation, enable or decouple methylation from a functional role, mean isolated CpG within TFBSs, etc. can still be important. By genomic analysis of six primates, the subset of non-polymorphic human-specific CpGs (~1.19 million), termed “CpG Beacons”, were identified (Bell et al. 2012). Significant clusters were found to be enriched for neurological and inflammatory disease-associated loci. It may be that these CpGs are enriched for epigenetic modifications in human-specific disease changes. The reduced statistical correction required with this limited set may enable the identification of novel associations and

could be used to test for potential increased or reduced human-specific regulatory effect in certain disease traits.

The dramatic rise in allergic disease indicates that immune pathways appear remarkably susceptible to modern environmental influences (Martino et al. 2011), that epigenetic insights into this rapid rise may be identified (Martino and Prescott 2012) and these may be human-specific. Human-specific CpG and DMR loci or similar approaches can be used to look for evidence for hypotheses such as this, as well as others such as early developmental effects on adult disease (Gluckman et al. 2008), inflammatory roles in obesity (Greenfield et al. 2004) or atherosclerosis (Varki et al. 2011; Libby et al. 2011).

13.9 Future Developments

Validation of the large number of potential disease-related methylation regions that may arise, particularly from a first stage array design, can be a practically difficult task. Now with the availability of non-biased targeted amplification techniques that can be used on BiS-converted DNA, such as Raindance (Tewhey et al. 2009), and subsequent coupling with second-generation sequencing, this will enable deep base-resolution targeted confirmation of these loci.

Once confirmed DMRs can be identified, functional assessment of these would previously be by the use of global epigenomic effectors, like demethylating agents, such as 5-azacytidine, in cell-line models. Precisely targeted assessment by methylation manipulation with artificial transcription factors (Rivenbark et al. 2012) or transcription activator-like effector nucleases TALENs (Reyon et al. 2012) are now being proposed and will permit more exact understanding of epigenetic change in these loci.

Base-resolution analysis of further modifications of DNA, for example hydroxymethylation (5hmC), has recently been shown to be possible with the chemical conversion technique, detailed by Booth et al., of selective chemical oxidation of 5hmC to 5fC with subsequent bisulphite conversion to uracil and then amplification to thymine (Booth et al. 2012). This pseudo-SNP generation then would be amenable to high-throughput array technology analysis, in the same fashion as for DNA methylation. Recently the possible importance of this 5hmC mark has been highlighted by an identified reduction across melanoma genomes (Lian et al. 2012), as well as increased levels within synaptic genes, exon–intron boundaries, and constitutively spliced exons within brain tissue, indicating a potential role in splicing in the central nervous system (Khare et al. 2012).

As mentioned the prospect of third-generation sequencing would be revolutionary for the field of epigenetics by directly enabling the assessment of DNA modifications without the necessity for DNA degradative chemical modifications (Flusberg et al. 2010; Clarke et al. 2009). Accurate detection of the low frequency modifications (5hmC, 5fC, 5caC and possibly other undiscovered) would allow these to be properly explored in more depth.

The clinical applications of DNA methylation profiling as a signature of human disease have particular promise (Heyn and Esteller 2012). As well as an additional dimension to integrate into genomic and transcriptomic data for precise pathological analysis (Bell and Beck 2009), further possibilities include minute detection of non-invasive cancer or signatures of inflammatory signals from surrogate biological fluids (Heyn and Esteller 2012). The high level of genetic mutation in epigenetic regulators and chromatin remodelling genes, in both cancer (Dawson et al. 2012) and developmental disorders (Kleefstra et al. 2006; Bienvenu and Chelly 2006), signal the importance of the accurate control of the epigenomic dimension in these complex disorders and that further dysregulation will be found. Finally, in the future with increasingly detailed base-resolution understanding of the epigenome, the prospect of “interventional epigenomics” may become possible. Due to the epigenome’s dynamic plasticity it may make more amenable to targeted therapeutic manipulation than static somatic mutation (Dawson and Kouzarides 2012).

References

- Aberg K, Khachane AN, Rudolf G, Nerella S, Fugman DA, Tischfield JA, van den Oord EJ (2012) Methylome-wide comparison of human genomic DNA extracted from whole blood and from EBV-transformed lymphocyte cell lines. *Eur J Hum Genet* 20(9):953–955. doi:[10.1038/ejhg.2012.33](https://doi.org/10.1038/ejhg.2012.33)
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2008) *Molecular biology of the cell*, 5th edn. Garland Science, New York [London: Taylor & Francis, distributor]
- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. *Nature* 437(7063):1299–1320
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265, doi:[10.1093/bioinformatics/bth457](https://doi.org/10.1093/bioinformatics/bth457) [pii]
- Bell CG (2011) Integration of genomic and epigenomic DNA methylation data in common complex diseases by haplotype-specific methylation analysis. *Pers Med* 8(3):243–251
- Bell CG, Beck S (2009) Advances in the identification and analysis of allele-specific expression. *Genome Med* 1(5):56, doi:[gm56](https://doi.org/10.1186/gm56) [pii][10.1186/gm56](https://doi.org/10.1186/gm56)
- Bell CG, Beck S (2010) The epigenomic interface between genome and environment in common complex diseases. *Brief Funct Genomics* 9(5–6):477–485. doi:[10.1093/bfgp/elq026](https://doi.org/10.1093/bfgp/elq026)
- Bell JT, Spector TD (2012) DNA methylation studies using twins: what are they telling us? *Genome Biol* 13(10):172. doi:[10.1186/gb-2012-13-10-172](https://doi.org/10.1186/gb-2012-13-10-172)
- Bell CG, Meyre D, Petretto E, Levy-Marchal C, Hercberg S, Charles MA, Boyle C, Weill J, Tauber M, Mein CA, Aitman TJ, Froguel P, Walley AJ (2007) No contribution of angiotensin-converting enzyme (ACE) gene variants to severe obesity: a model for comprehensive case/control and quantitative cladistic analysis of ACE in human diseases. *Eur J Hum Genet* 15(3):320–327. doi:[10.1038/sj.ejhg.5201754](https://doi.org/10.1038/sj.ejhg.5201754)
- Bell CG, Teschendorff AE, Rakyan VK, Maxwell AP, Beck S, Savage DA (2010a) Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Med Genomics* 3:33, doi:[1755-8794-3-33](https://doi.org/10.1186/1755-8794-3-33) [pii][10.1186/1755-8794-3-33](https://doi.org/10.1186/1755-8794-3-33)
- Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, Akan P, Stupka E, Down TA, Prokopenko I, Morison IM, Mill J, Pidsley R, Deloukas P, Frayling TM, Hattersley AT, McCarthy MI, Beck S, Hitman GA (2010b) Integrated genetic and epigenetic analysis

- identifies haplotype-specific methylation in the FTO Type 2 diabetes and obesity susceptibility locus. *PLoS One* 5(11):e14040. doi:[10.1371/journal.pone.0014040](https://doi.org/10.1371/journal.pone.0014040)
- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12(1):R10. doi:[10.1186/gb-2011-12-1-r10](https://doi.org/10.1186/gb-2011-12-1-r10)
- Bell CG, Wilson GA, Butcher LM, Roos C, Walter L, Beck S (2012) Human-specific CpG “beacons” identify loci associated with human-specific traits and disease. *Epigenetics* 7(10):1188–1199. doi:[10.4161/epi.22127](https://doi.org/10.4161/epi.22127)
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98(4):288–295. doi:[10.1016/j.ygeno.2011.07.007](https://doi.org/10.1016/j.ygeno.2011.07.007)
- Bienvenu T, Chelly J (2006) Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat Rev Genet* 7(6):415–426. doi:[nrg1878](https://doi.org/10.1038/nrg1878) [pii][10.1038/nrg1878](https://doi.org/10.1038/nrg1878)
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16(1):6–21. doi:[10.1101/gad.947102](https://doi.org/10.1101/gad.947102)
- Bird A (2011) The dinucleotide CG as a genomic signalling module. *J Mol Biol* 409(1):47–53. doi:[10.1016/j.jmb.2011.01.056](https://doi.org/10.1016/j.jmb.2011.01.056)
- Birney E (2011) Chromatin and heritability: how epigenetic studies can complement genetic approaches. *Trends Genet* 27(5):172–176. doi:[S0168-9525\(11\)00031-X](https://doi.org/10.1016/j.tig.2011.02.005) [pii][10.1016/j.tig.2011.02.005](https://doi.org/10.1016/j.tig.2011.02.005)
- Blackledge NP, Klose R (2011) CpG island chromatin: a platform for gene regulation. *Epigenetics* 6(2):147–152. doi:[10.1016/j.epi.2011.01.009](https://doi.org/10.1016/j.epi.2011.01.009)
- Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol Cell* 38(2):179–190. doi:[S1097-2765\(10\)00286-8](https://doi.org/10.1016/j.molcel.2010.04.009) [pii][10.1016/j.molcel.2010.04.009](https://doi.org/10.1016/j.molcel.2010.04.009)
- Bock C (2012) Analysing and interpreting DNA methylation data. *Nat Rev Genet* 13(10):705–719. doi:[10.1038/nrg3273](https://doi.org/10.1038/nrg3273)
- Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A (2012) DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell* 47(4):633–647. doi:[10.1016/j.molcel.2012.06.019](https://doi.org/10.1016/j.molcel.2012.06.019)
- Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, Balasubramanian S (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336(6083):934–937. doi:[10.1126/science.1220671](https://doi.org/10.1126/science.1220671)
- Branco MR, Ficz G, Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet* 13(1):7–13. doi:[10.1038/nrg3080](https://doi.org/10.1038/nrg3080)
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 88(4):450–457. doi:[S0002-9297\(11\)00095-4](https://doi.org/10.1016/j.ajhg.2011.03.003) [pii][10.1016/j.ajhg.2011.03.003](https://doi.org/10.1016/j.ajhg.2011.03.003)
- Brena RM, Huang TH, Plass C (2006) Toward a human epigenome. *Nat Genet* 38(12):1359–1360. doi:[ng1206-1359](https://doi.org/10.1038/ng1206-1359) [pii][10.1038/ng1206-1359](https://doi.org/10.1038/ng1206-1359)
- Brennan EP, Ehrlich M, Brazil DP, Crean JK, Murphy M, Sadlier DM, Martin F, Godson C, McKnight AJ, van den Boom D, Maxwell AP, Savage DA (2009) Comparative analysis of DNA methylation profiles in peripheral blood leukocytes versus lymphoblastoid cell lines. *Epigenetics* 4(3):159–164
- Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, Herwig R, Adjaye J (2010) Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. *Genome Res* 20(10):1441–1450. doi:[10.1101/gr.110114.110](https://doi.org/10.1101/gr.110114.110)
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4(4):265–270. doi:[nnano.2009.12](https://doi.org/10.1038/nnano.2009.12) [pii][10.1038/nnano.2009.12](https://doi.org/10.1038/nnano.2009.12)
- Clarke PM, Glasziou P, Patel A, Chalmers J, Woodward M, Harrap SB, Salomon JA (2010) Event rates, hospital utilization, and costs associated with major complications of diabetes: a

- multicountry comparative analysis. *PLoS Med* 7(2):e1000236. doi:[10.1371/journal.pmed.1000236](https://doi.org/10.1371/journal.pmed.1000236)
- Conerly ML, Teves SS, Diolaiti D, Ulrich M, Eisenman RN, Henikoff S (2010) Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res* 20 (10):1383–1390. doi:[10.1101/gr.106542.110](https://doi.org/10.1101/gr.106542.110)
- Cooper DN, Mort M, Stenson PD, Ball EV, Chuzhanova NA (2010) Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum Genomics* 4 (6):406–410
- Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, Al-Sarraj S, Dobson R, Schalkwyk LC, Mill J (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol* 13(6):R43. doi:[10.1186/gb-2012-13-6-r43](https://doi.org/10.1186/gb-2012-13-6-r43)
- Dawson MA, Kouzarides T (2012) Cancer epigenetics: from mechanism to therapy. *Cell* 150 (1):12–27. doi:[10.1016/j.cell.2012.06.013](https://doi.org/10.1016/j.cell.2012.06.013)
- Dawson MA, Kouzarides T (2012) Targeting epigenetic readers in cancer. *N Engl J Med* 367(7):647–657. doi:[10.1056/NEJMr1112635](https://doi.org/10.1056/NEJMr1112635)
- De S, Babu MM (2010) A time-invariant principle of genome evolution. *Proc Natl Acad Sci U S A* 107(29):13004–13009. doi:[0914454107](https://doi.org/10.1073/pnas.0914454107) [pii][10.1073/pnas.0914454107](https://doi.org/10.1073/pnas.0914454107)
- Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F (2011) Evaluation of the Infinium methylation 450K technology. *Epigenomics* 3(6):771–784. doi:[10.2217/epi.11.105](https://doi.org/10.2217/epi.11.105)
- Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Hercberg S, Le Stunff C, Bougneres P, Kovacs P, Marre M, Balkau B, Cauchi S, Chevre JC, Froguel P (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 39(6):724–726. doi:[ng2048](https://doi.org/10.1038/ng2048) [pii][10.1038/ng2048](https://doi.org/10.1038/ng2048)
- Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41(12):1350–1353. doi:[ng.471](https://doi.org/10.1038/ng.471) [pii][10.1038/ng.471](https://doi.org/10.1038/ng.471)
- Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, Sorek R, Rechavi G (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485 (7397):201–206. doi:[10.1038/nature11112](https://doi.org/10.1038/nature11112)
- Duncan BK, Miller JH (1980) Mutagenic deamination of cytosine residues in DNA. *Nature* 287 (5782):560–561
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38 (12):1378–1385. doi:[ng1909](https://doi.org/10.1038/ng1909) [pii][10.1038/ng1909](https://doi.org/10.1038/ng1909)
- Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9(3):215–216. doi:[10.1038/nmeth.1906](https://doi.org/10.1038/nmeth.1906)
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43–49. doi:[10.1038/nature09906](https://doi.org/10.1038/nature09906)
- Esteller M (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8(4):286–298. doi:[nrg2005](https://doi.org/10.1038/nrg2005) [pii][10.1038/nrg2005](https://doi.org/10.1038/nrg2005)
- Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD (2012) Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A* 109(19):7332–7337. doi:[10.1073/pnas.1201310109](https://doi.org/10.1073/pnas.1201310109)

- Faulk C, Dolinoy DC (2011) Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. *Epigenetics* 6(7)
- Feil R, Fraga MF (2011) Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13(2):97–109. doi:[10.1038/nrg3142](https://doi.org/10.1038/nrg3142)
- Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447(7143):433–440. doi:[nature05919](https://doi.org/10.1038/nature05919) [pii][10.1038/nature05919](https://doi.org/10.1038/nature05919)
- Feinberg AP (2008) Epigenetics at the epicenter of modern medicine. *JAMA* 299(11):1345–1350. doi:[299/11/1345](https://doi.org/10.1001/jama.299.11.1345) [pii][10.1001/jama.299.11.1345](https://doi.org/10.1001/jama.299.11.1345)
- Feinberg AP, Irizarry RA, Fradin D, Aryee MJ, Murakami P, Aspelund T, Eiriksdottir G, Harris TB, Launer L, Gudnason V, Fallin MD (2010) Personalized epigenomic signatures that are stable over time and covary with body mass index. *Sci Transl Med* 2(49):49–67. doi:[2/49/49ra67](https://doi.org/10.1126/scitranslmed.3001262) [pii][10.1126/scitranslmed.3001262](https://doi.org/10.1126/scitranslmed.3001262)
- Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* 12(8):565–575. doi:[10.1038/nrg3032](https://doi.org/10.1038/nrg3032)
- Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O, Ferrer I, Sanchez-Cespedes M, Villanueva A, Carmona J, Sanchez-Mut JV, Berdasco M, Moreno V, Capella G, Monk D, Ballestar E, Ropero S, Martinez R, Sanchez-Carbayo M, Prosper F, Agirre X, Fraga MF, Grana O, Perez-Jurado L, Mora J, Puig S, Prat J, Badimon L, Puca AA, Meltzer SJ, Lengauer T, Bridgewater J, Bock C, Esteller M (2012) A DNA methylation fingerprint of 1628 human samples. *Genome Res* 22(2):407–419. doi:[10.1101/gr.119867.110](https://doi.org/10.1101/gr.119867.110)
- Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7(6):461–465. doi:[nmeth.1459](https://doi.org/10.1038/nmeth.1459) [pii][10.1038/nmeth.1459](https://doi.org/10.1038/nmeth.1459)
- Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316(5826):889–894. doi:[1141634](https://doi.org/10.1126/science.1141634) [pii][10.1126/science.1141634](https://doi.org/10.1126/science.1141634)
- Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6(5):e1000952. doi:[10.1371/journal.pgen.1000952](https://doi.org/10.1371/journal.pgen.1000952)
- Gluckman PD, Hanson MA, Cooper C, Thornburg KL (2008) Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med* 359(1):61–73. doi:[359/1/61](https://doi.org/10.1056/NEJMr0708473) [pii][10.1056/NEJMr0708473](https://doi.org/10.1056/NEJMr0708473)
- Greenfield JR, Samaras K, Jenkins AB, Kelly PJ, Spector TD, Gallimore JR, Pepys MB, Campbell LV (2004) Obesity is an important determinant of baseline serum C-reactive protein concentration in monozygotic twins, independent of genetic influences. *Circulation* 109(24):3022–3028. doi:[10.1161/01.CIR.0000130640.77501.79](https://doi.org/10.1161/01.CIR.0000130640.77501.79)
- Hansen KD, Timp W, Bravo HC, Sabuncuyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet.* doi:[10.1038/ng.865](https://doi.org/10.1038/ng.865)
- Heijmans BT, Mill J (2012) Commentary: the seven plagues of epigenetic epidemiology. *Int J Epidemiol* 41(1):74–78. doi:[10.1093/ije/dyr225](https://doi.org/10.1093/ije/dyr225)
- Heijmans BT, Tobin EW, Lumey LH, Slagboom PE (2009) The epigenome: archive of the prenatal environment. *Epigenetics* 4(8):526–531
- Hellman A, Chess A (2007) Gene body-specific methylation on the active X chromosome. *Science* 315(5815):1141–1143. doi:[315/5815/1141](https://doi.org/10.1126/science.1136352) [pii][10.1126/science.1136352](https://doi.org/10.1126/science.1136352)

- Heyn H, Esteller M (2012) DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* 13(10):679–692. doi:[10.1038/nrg3270](https://doi.org/10.1038/nrg3270)
- Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ, Puca AA, Sayols S, Pujana MA, Serra-Musach J, Iglesias-Platas I, Formiga F, Fernandez AF, Fraga MF, Heath SC, Valencia A, Gut IG, Wang J, Esteller M (2012) Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* 109(26):10522–10527. doi:[10.1073/pnas.1120658109](https://doi.org/10.1073/pnas.1120658109)
- Hindorf LA, Sethupathy P, Jenkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106(23):9362–9367. doi:[10.1073/pnas.0903103106](https://doi.org/10.1073/pnas.0903103106) [pii]
- Hodges E, Molaro A, Dos Santos CO, Thekkat P, Song Q, Uren PJ, Park J, Butler J, Rafii S, McCombie WR, Smith AD, Hannon GJ (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* 44(1):17–28. doi:[10.1016/j.molcel.2011.08.026](https://doi.org/10.1016/j.molcel.2011.08.026)
- Houssman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13(1):86. doi:[10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86)
- Illingworth RS, Bird AP (2009) CpG islands—‘a rough guide’. *FEBS Lett* 583(11):1713–1720. doi:[10.1016/j.febslet.2009.04.012](https://doi.org/10.1016/j.febslet.2009.04.012)
- Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* 6(1):e22. doi:[10.1371/journal.pbio.0060022](https://doi.org/10.1371/journal.pbio.0060022)
- Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* 6(9). doi:[10.1371/journal.pgen.1001134](https://doi.org/10.1371/journal.pgen.1001134)
- Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, Ji H, Potash JB, Sabuncian S, Feinberg AP (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41(2):178–186. doi:[ng.298](https://doi.org/10.1038/ng.298) [pii][10.1038/ng.298](https://doi.org/10.1038/ng.298)
- Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA (2012) Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 41(1):200–209. doi:[10.1093/ije/dyr238](https://doi.org/10.1093/ije/dyr238)
- Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O’Hanlon TP, Rider LG, Jacinto FV, Lopez-Longo FJ, Dopazo J, Forn M, Peinado MA, Carreno L, Sawalha AH, Harley JB, Siebert R, Esteller M, Miller FW, Ballestar E (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 20(2):170–179. doi:[10.1101/gr.100289.109](https://doi.org/10.1101/gr.100289.109) [pii]
- Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, Yi C, Lindahl T, Pan T, Yang YG, He C (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* 7(12):885–887. doi:[10.1038/nchembio.687](https://doi.org/10.1038/nchembio.687)
- Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10(3):161–172. doi:[10.1038/nrg2522](https://doi.org/10.1038/nrg2522)
- Johnson MD, Mueller M, Game L, Aitman TJ (2012) Single nucleotide analysis of cytosine methylation by whole-genome shotgun bisulfite sequencing. *Curr Protoc Mol Biol* Chapter 21:Unit21 23. doi:[10.1002/0471142727.mb2123s99](https://doi.org/10.1002/0471142727.mb2123s99)
- Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10(11):805–811. doi:[10.1038/nrg2651](https://doi.org/10.1038/nrg2651)
- Jonsson T, Atwal JK, Steinberg S, Snaedal J, Jonsson PV, Bjornsson S, Stefansson H, Sulem P, Gudbjartsson D, Maloney J, Hoyte K, Gustafson A, Liu Y, Lu Y, Bhargale T, Graham RR, Huttenlocher J, Bjornsdottir G, Andreassen OA, Jonsson EG, Palotie A, Behrens TW,

- Magnusson OT, Kong A, Thorsteinsdottir U, Watts RJ, Stefansson K (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488 (7409):96–99. doi:[10.1038/nature11283](https://doi.org/10.1038/nature11283)
- Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, Huang Z, Hoyo C, Midttun O, Cupul-Uicab LA, Ueland PM, Wu MC, Nystad W, Bell DA, Peddada SD, London SJ (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120(10):1425–1431. doi:[10.1289/ehp.1205412](https://doi.org/10.1289/ehp.1205412)
- Kaikkonen MU, Lam MT, Glass CK (2011) Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res* 90(3):430–440. doi:[10.1093/cvr/cvr097](https://doi.org/10.1093/cvr/cvr097)
- Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 40 (7):904–908. doi:[ng.174](https://doi.org/10.1038/ng.174) [pii][10.1038/ng.174](https://doi.org/10.1038/ng.174)
- Khare T, Pai S, Koncevicius K, Pal M, Kriukiene E, Liutkeviciute Z, Irimia M, Jia P, Ptak C, Xia M, Tice R, Tochigi M, Morera S, Nazarians A, Belsham D, Wong AH, Blencowe BJ, Wang SC, Kapranov P, Kustra R, Labrie V, Klimasauskas S, Petronis A (2012) 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat Struct Mol Biol*. doi:[10.1038/nsmb.2372](https://doi.org/10.1038/nsmb.2372)
- Kilpelainen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, Ahmad T, Mora S, Kaakinen M, Sandholt CH, Holzapfel C, Autenrieth CS, Hypponen E, Cauchi S, He M, Kutalik Z, Kumari M, Stancakova A, Meidtner K, Balkau B, Tan JT, Mangino M, Timpson NJ, Song Y, Zillikens MC, Jablonski KA, Garcia ME, Johansson S, Bragg-Gresham JL, Wu Y, van Vliet-Ostapchouk JV, Onland-Moret NC, Zimmermann E, Rivera NV, Tanaka T, Stringham HM, Silbernagel G, Kanoni S, Feitosa MF, Snitker S, Ruiz JR, Metter J, Larrad MT, Atalay M, Hakanen M, Amin N, Cavalcanti-Proenca C, Grontved A, Hallmans G, Jansson JO, Kuusisto J, Kahonen M, Lutsey PL, Nolan JJ, Palla L, Pedersen O, Perusse L, Renstrom F, Scott RA, Shungin D, Sovio U, Tammelin TH, Ronnema T, Lakka TA, Uusitupa M, Rios MS, Ferrucci L, Bouchard C, Meirhaeghe A, Fu M, Walker M, Borecki IB, Dedoussis GV, Fritsche A, Ohlsson C, Boehnke M, Bandinelli S, van Duijn CM, Ebrahim S, Lawlor DA, Gudnason V, Harris TB, Sorensen TI, Mohlke KL, Hofman A, Uitterlinden AG, Tuomilehto J, Lehtimäki T, Raitakari O, Isomaa B, Njolstad PR, Florez JC, Liu S, Ness A, Spector TD, Tai ES, Froguel P, Boeing H, Laakso M, Marmot M, Bergmann S, Power C, Khaw KT, Chasman D, Ridker P, Hansen T, Monda KL, Illig T, Jarvelin MR, Wareham NJ, Hu FB, Groop LC, Orho-Melander M, Ekelund U, Franks PW, Loos RJ (2011) Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med* 8(11):e1001116. doi:[10.1371/journal.pmed.1001116](https://doi.org/10.1371/journal.pmed.1001116)
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465(7295):182–187. doi:[nature09033](https://doi.org/10.1038/nature09033) [pii][10.1038/nature09033](https://doi.org/10.1038/nature09033)
- Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, Genevieve D, Cormier-Daire V, van Esch H, Fryns JP, Hamel BC, Sistermans EA, de Vries BB, van Bokhoven H (2006) Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. *Am J Hum Genet* 79(2):370–377. doi:[S0002-9297\(07\)63145-0](https://doi.org/10.1086/505693) [pii][10.1086/505693](https://doi.org/10.1086/505693)
- Li JB, Gao Y, Aach J, Zhang K, Kryukov GV, Xie B, Ahlford A, Yoon JK, Rosenbaum AM, Zaranek AW, LeProust E, Sunyaev SR, Church GM (2009) Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res* 19(9):1606–1615. doi:[10.1101/gr.092213.109](https://doi.org/10.1101/gr.092213.109)
- Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, Zheng H, Yu J, Wu H, Sun J, Zhang H, Chen Q, Luo R, Chen M, He Y, Jin X, Zhang Q, Yu C, Zhou G, Huang Y, Cao H, Zhou X, Guo S, Hu X, Li X, Kristiansen K, Bolund L, Xu J, Wang W, Yang H, Wang J, Li R, Beck S, Zhang X (2010) The

- DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol* 8(11):e1000533. doi:[10.1371/journal.pbio.1000533](https://doi.org/10.1371/journal.pbio.1000533)
- Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, Lee CW, Hu D, Lian BQ, Kleffel S, Yang Y, Neiswender J, Khorasani AJ, Fang R, Lezcano C, Duncan LM, Scolyer RA, Thompson JF, Kakavand H, Houvras Y, Zon LI, Mihm MC Jr, Kaiser UB, Schatton T, Woda BA, Murphy GF, Shi YG (2012) Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150(6):1135–1146. doi:[10.1016/j.cell.2012.07.033](https://doi.org/10.1016/j.cell.2012.07.033)
- Libby P, Ridker PM, Hansson GK (2011) Progress and challenges in translating the biology of atherosclerosis. *Nature* 473(7347):317–325
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43(11):1091–1097. doi:[10.1038/ng.946](https://doi.org/10.1038/ng.946)
- Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alfoldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Baldwin J, Bloom T, Chin CW, Heiman D, Nicol R, Nusbaum C, Young S, Wilkinson J, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Cree A, Dihm HH, Fowler G, Jhangiani S, Joshi V, Lee S, Lewis LR, Nazareth LV, Okwuonu G, Santibanez J, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Delehaunty K, Dooling D, Fronik C, Fulton L, Fulton B, Graves T, Minx P, Sodergren E, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370):476–482. doi:[10.1038/nature10530](https://doi.org/10.1038/nature10530)
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322. doi:[nature08514](https://doi.org/10.1038/nature08514) [pii][10.1038/nature08514](https://doi.org/10.1038/nature08514)
- Liu Y, Siegmund KD, Laird PW, Berman BP (2012) Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol* 13(7):R61. doi:[10.1186/gb-2012-13-7-r61](https://doi.org/10.1186/gb-2012-13-7-r61)
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753. doi:[10.1038/nature08494](https://doi.org/10.1038/nature08494)
- Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* 11(4):285–296. doi:[nrg2752](https://doi.org/10.1038/nrg2752) [pii][10.1038/nrg2752](https://doi.org/10.1038/nrg2752)
- Martin DI, Singer M, Dhahbi J, Mao G, Zhang L, Schroth GP, Pachter L, Boffelli D (2011) Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. *Genome Res*. doi:[10.1101/gr.122721.111](https://doi.org/10.1101/gr.122721.111)
- Martino DJ, Prescott SL (2012) Progress in understanding the epigenetic basis for immune development, immune function, and the rising incidence of allergic disease. *Curr Allergy Asthma Rep*. doi:[10.1007/s11882-012-0312-1](https://doi.org/10.1007/s11882-012-0312-1)
- Martino DJ, Tulic MK, Gordon L, Hodder M, Richman TR, Metcalfe J, Prescott SL, Saffery R (2011) Evidence for age-related and individual-specific changes in DNA methylation profile of mononuclear cells during early immune development in humans. *Epigenetics* 6(9):1085–1094. doi:[10.4161/epi.6.9.16401](https://doi.org/10.4161/epi.6.9.16401)
- Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in

- regulating alternative promoters. *Nature* 466(7303):253–257, doi:nature09165 [pii][10.1038/nature09165](https://doi.org/10.1038/nature09165)
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayvin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–1195. doi:[10.1126/science.1222794](https://doi.org/10.1126/science.1222794)
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 9(5):356–369, doi:nrg2344 [pii][10.1038/nrg2344](https://doi.org/10.1038/nrg2344)
- Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149(7):1635–1646. doi:[10.1016/j.cell.2012.05.003](https://doi.org/10.1016/j.cell.2012.05.003)
- Mill J, Tang T, Kaminsky Z, Khare T, Yazdanpanah S, Bouchard L, Jia P, Assadzadeh A, Flanagan J, Schumacher A, Wang SC, Petronis A (2008) Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *Am J Hum Genet* 82(3):696–711, doi:S0002-9297(08)00148-1 [pii][10.1016/j.ajhg.2008.01.008](https://doi.org/10.1016/j.ajhg.2008.01.008)
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146(6):1029–1041. doi:[10.1016/j.cell.2011.08.016](https://doi.org/10.1016/j.cell.2011.08.016)
- Monick MM, Beach SR, Plume J, Sears R, Gerrard M, Brody GH, Philibert RA (2012) Coordinated changes in AHRR methylation in lymphoblasts and pulmonary macrophages from smokers. *Am J Med Genet B Neuropsychiatr Genet* 159B(2):141–151. doi:[10.1002/ajmg.b.32021](https://doi.org/10.1002/ajmg.b.32021)
- Murrell A, Rakyan VK, Beck S (2005) From genome to epigenome. *Hum Mol Genet* 14 Spec No 1:R3–R10. doi:14/suppl_1/R3 [pii][10.1093/hmg/ddi110](https://doi.org/10.1093/hmg/ddi110)
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kutayvin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA (2012a) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489(7414):83–90. doi:[10.1038/nature11212](https://doi.org/10.1038/nature11212)
- Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA (2012b) Circuitry and dynamics of human transcription factor regulatory networks. *Cell*. doi:[10.1016/j.cell.2012.04.040](https://doi.org/10.1016/j.cell.2012.04.040)
- Nguyen A, Rauch TA, Pfeifer GP, Hu VW (2010) Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB J* 24(8):3036–3051. doi:[10.1096/fj.10-154484](https://doi.org/10.1096/fj.10-154484)
- Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y (2011) A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet* 7(2):e1001316. doi:[10.1371/journal.pgen.1001316](https://doi.org/10.1371/journal.pgen.1001316)
- Philibert RA, Beach SR, Brody GH (2012) Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. *Epigenetics* 7(11)
- Qi Q, Chu AY, Kang JH, Jensen MK, Curhan GC, Pasquale LR, Ridker PM, Hunter DJ, Willett WC, Rimm EB, Chasman DI, Hu FB, Qi L (2012) Sugar-sweetened beverages and genetic risk of obesity. *N Engl J Med*. doi:[10.1056/NEJMoa1203039](https://doi.org/10.1056/NEJMoa1203039)
- Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20(4):434–439, doi:gr.103101.109 [pii][10.1101/gr.103101.109](https://doi.org/10.1101/gr.103101.109)

- Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12(8):529–541. doi:[10.1038/nrg3000](https://doi.org/10.1038/nrg3000)
- Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30(5):460–465. doi:[10.1038/nbt.2170](https://doi.org/10.1038/nbt.2170)
- Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nat Rev Genet* 7(5):395–401. doi:[nrg1834](https://doi.org/10.1038/nrg1834) [pii][10.1038/nrg1834](https://doi.org/10.1038/nrg1834)
- Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, Blancafort P (2012) Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* 7(4):350–360. doi:[10.4161/epi.19507](https://doi.org/10.4161/epi.19507)
- Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328(5978):636–639. doi:[science.1186802](https://doi.org/10.1126/science.1186802) [pii][10.1126/science.1186802](https://doi.org/10.1126/science.1186802)
- Robertson KD (2005) DNA methylation and human disease. *Nat Rev Genet* 6(8):597–610
- Robinson MD, Statham AL, Speed TP, Clark SJ (2010) Protocol matters: which methylome are you actually studying? *Epigenomics* 2(4):587–598. doi:[10.2217/epi.10.36](https://doi.org/10.2217/epi.10.36)
- Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, Plomin R, Mill J (2010) Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 86(2):196–212. doi:[S0002-9297\(10\)00017-0](https://doi.org/10.1016/j.ajhg.2010.01.014) [pii][10.1016/j.ajhg.2010.01.014](https://doi.org/10.1016/j.ajhg.2010.01.014)
- Schmidt D, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT (2010) Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328(5981):1036–1040. doi:[science.1186176](https://doi.org/10.1126/science.1186176) [pii][10.1126/science.1186176](https://doi.org/10.1126/science.1186176)
- Schones DE, Zhao K (2008) Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* 9(3):179–191. doi:[10.1038/nrg2270](https://doi.org/10.1038/nrg2270)
- Shoemaker R, Deng J, Wang W, Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* 20(7):883–889. doi:[gr.104695.109](https://doi.org/10.1101/gr.104695.109) [pii][10.1101/gr.104695.109](https://doi.org/10.1101/gr.104695.109)
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371):74–79. doi:[10.1038/nature10442](https://doi.org/10.1038/nature10442)
- Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H (1999) DNA methylation represses transcription in vivo. *Nat Genet* 22(2):203–206. doi:[10.1038/9727](https://doi.org/10.1038/9727)
- Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445(7130):881–885
- Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics* 182(3):845–850. doi:[10.1534/genetics.109.102798](https://doi.org/10.1534/genetics.109.102798)
- Spitz F, Furlong EE (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13(9):613–626. doi:[10.1038/nrg3207](https://doi.org/10.1038/nrg3207)
- Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, Suter CM, Preiss T (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res* 40(11):5023–5033. doi:[10.1093/nar/gks144](https://doi.org/10.1093/nar/gks144)
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480(7378):490–495. doi:[10.1038/nature10716](https://doi.org/10.1038/nature10716)
- Strachan T, Read AP (1999) Human molecular genetics, 2nd edn. Wiley, New York
- Su J, Yan H, Wei Y, Liu H, Wang F, Lv J, Wu Q, Zhang Y (2012) CpG_MPs: identification of CpG methylation patterns of genomic regions from high-throughput bisulfite sequencing data. *Nucleic Acids Res*. doi:[10.1093/nar/gks829](https://doi.org/10.1093/nar/gks829)

- Tan M, Luo H, Lee S, Jin F, Yang JS, Montellier E, Buchou T, Cheng Z, Rousseaux S, Rajagopal N, Lu Z, Ye Z, Zhu Q, Wysocka J, Ye Y, Khochbin S, Ren B, Zhao Y (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146(6):1016–1028. doi:[10.1016/j.cell.2011.08.008](https://doi.org/10.1016/j.cell.2011.08.008)
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ, Widschwendter M (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 20(4):440–446. doi:[gr.103606.109](https://doi.org/gr.103606.109) [pii][10.1101/gr.103606.109](https://doi.org/10.1101/gr.103606.109)
- Tewhey R, Warner JB, Nakano M, Libby B, Medkova M, David PH, Kotsopoulos SK, Samuels ML, Hutchison JB, Larson JW, Topol EJ, Weiner MP, Harismendy O, Olson J, Link DR, Frazer KA (2009) Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat Biotechnol* 27(11):1025–1031. doi:[10.1038/nbt.1583](https://doi.org/10.1038/nbt.1583)
- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661–678. doi:[nature05911](https://doi.org/nature05911) [pii][10.1038/nature05911](https://doi.org/10.1038/nature05911)
- Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464(7291):1082–1086. doi:[nature08924](https://doi.org/nature08924) [pii][10.1038/nature08924](https://doi.org/10.1038/nature08924)
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, Diegel M, Dunn D, Ebersol AK, Frum T, Giste E, Johnson AK, Johnson EM, Kutayavin T, Lajoie B, Lee BK, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen ED, Qu H, Reynolds AP, Roach V, Safi A, Sanchez ME, Sanyal A, Shafer A, Simon JM, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Lenhard B, Tewari M, Dorschner MO, Hansen RS, Navas PA, Stamatoyannopoulos G, Iyer VR, Lieb JD, Sunyaev SR, Akey JM, Sabo PJ, Kaul R, Furey TS, Dekker J, Crawford GE, Stamatoyannopoulos JA (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82. doi:[10.1038/nature11232](https://doi.org/10.1038/nature11232)
- Tycko B (2010) Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet* 19(R2):R210–R220. doi:[10.1093/hmg/ddq376](https://doi.org/10.1093/hmg/ddq376)
- Varki A, Geschwind DH, Eichler EE (2008) Explaining human uniqueness: genome interactions with environment, behaviour and culture. *Nat Rev Genet* 9(10):749–763. doi:[nrg2428](https://doi.org/nrg2428) [pii][10.1038/nrg2428](https://doi.org/10.1038/nrg2428)
- Varki NM, Strobert E, Dick EJ Jr, Benirschke K, Varki A (2011) Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology. *Annu Rev Pathol* 6:365–393. doi:[10.1146/annurev-pathol-011110-130315](https://doi.org/10.1146/annurev-pathol-011110-130315)
- Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M, Sandstrom R, Thurman RE, Kaul R, Myers RM, Stamatoyannopoulos JA (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* 22(9):1680–1688. doi:[10.1101/gr.136101.111](https://doi.org/10.1101/gr.136101.111)
- Ward LD, Kellis M (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 40 (Database issue):7
- Waterland RA, Kellermayer R, Laritsky E, Rayco-Solon P, Harris RA, Travisano M, Zhang W, Torskaya MS, Zhang J, Shen L, Manary MJ, Prentice AM (2010) Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet* 6(12):e1001252
- Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39(4):457–466. doi:[ng1990](https://doi.org/ng1990) [pii][10.1038/ng1990](https://doi.org/10.1038/ng1990)
- Wilson G, Dhami P, Feber A, Cortazar D, Suzuki Y, Schulz R, Schar P, Beck S (2012) Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers. *GigaScience* 1(1):3

- Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Magi R, Waite L, Vernon Smith A, Yerges-Armstrong LM, Monda KL, Hadley D, Mahajan A, Li G, Kapur K, Vitart V, Huffman JE, Wang SR, Palmer C, Esko T, Fischer K, Hua Zhao J, Demirkan A, Isaacs A, Feitosa MF, Luan J, Heard-Costa NL, White C, Jackson AU, Preuss M, Ziegler A, Eriksson J, Kutalik Z, Frau F, Nolte IM, Van Vliet-Ostaptchouk JV, Hottenga JJ, Jacobs KB, Verweij N, Goel A, Medina-Gomez C, Estrada K, Lynn Bragg-Gresham J, Sanna S, Sidore C, Tyrer J, Teumer A, Prokopenko I, Mangino M, Lindgren CM, Assimes TL, Shuldiner AR, Hui J, Beilby JP, McArdle WL, Hall P, Haritunians T, Zgaga L, Kolcic I, Polasek O, Zemunik T, Oostra BA, Juhani Juntila M, Gronberg H, Schreiber S, Peters A, Hicks AA, Stephens J, Foad NS, Laitinen J, Pouta A, Kaakinen M, Willemsen G, Vink JM, Wild SH, Navis G, Asselbergs FW, Homuth G, John U, Iribarren C, Harris T, Launer L, Gudnason V, O'Connell JR, Boerwinkle E, Cadby G, Palmer LJ, James AL, Musk AW, Ingelsson E, Psaty BM, Beckmann JS, Waeber G, Vollenweider P, Hayward C, Wright AF, Rudan I, Groop LC, Metspalu A, Tee Khaw K, van Duijn CM, Borecki IB, Province MA, Wareham NJ, Tardif JC, Huikuri HV, Adrienne Cupples L, Atwood LD, Fox CS, Boehnke M, Collins FS, Mohlke KL, Erdmann J, Schunkert H, Hengstenberg C, Stark K, Lorentzon M, Ohlsson C, Cusi D, Staessen JA, Van der Klauw MM, Pramstaller PP, Kathiresan S, Jolley JD, Ripatti S, Jarvelin MR, de Geus EJ, Boomsma DI, Penninx B, Wilson JF, Campbell H, Chanock SJ, van der Harst P, Hamsten A, Watkins H, Hofman A, Witteman JC, Carola Zillikens M, Uitterlinden AG, Rivadeneira F, Kiemeny LA, Vermeulen SH, Abecasis GR, Schlessinger D, Schipf S, Stumvoll M, Tonjes A, Spector TD, North KE, Lettre G, McCarthy MI, Berndt SI, Heath AC, Madden PA, Nyholt DR, Montgomery GW, Martin NG, McKnight B, Strachan DP, Hill WG, Snieder H, Ridker PM, Thorsteinsdottir U, Stefansson K, Frayling TM, Hirschhorn JN, Goddard ME, Visscher PM (2012) FTO genotype is associated with phenotypic variability of body mass index. *Nature*. doi:[10.1038/nature11401](https://doi.org/10.1038/nature11401)
- Zeng J, Konopka G, Hunt BG, Preuss TM, Geschwind D, Yi SV (2012) Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am J Hum Genet*. doi:[10.1016/j.ajhg.2012.07.024](https://doi.org/10.1016/j.ajhg.2012.07.024)
- Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, Craig DW, Redman M, Gershon ES, Liu C (2010) Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet* 86(3):411–419, doi:S0002-9297(10)00087-X [pii][10.1016/j.ajhg.2010.02.005](https://doi.org/10.1016/j.ajhg.2010.02.005)
- Zhou VW, Goren A, Bernstein BE (2011a) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12(1):7–18. doi:[10.1038/nrg2905](https://doi.org/10.1038/nrg2905)
- Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, Koebbe BC, Nielsen C, Hirst M, Farnham P, Kuhn RM, Zhu J, Smirnov I, Kent WJ, Haussler D, Madden PA, Costello JF, Wang T (2011b) The human epigenome browser at Washington University. *Nat Methods* 8(12):989–990. doi:[10.1038/nmeth.1772](https://doi.org/10.1038/nmeth.1772)
- Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 109(4):1193–1198. doi:[10.1073/pnas.1119675109](https://doi.org/10.1073/pnas.1119675109)

Chapter 14

Analytical Considerations for Epigenome-Wide Association Scans of Complex Traits

Jordana T. Bell

Abstract Recent advances in molecular methods and next-generation sequencing technologies have allowed for the detection of epigenetic variation at an unprecedented level of resolution. With the availability of new genome-wide epigenetic assays, particularly for DNA methylation, epigenetic studies of human complex traits have moved towards epigenome-wide association scans (EWAS). Similar to genome-wide association scans (GWAS), EWAS aim to perform a genome-wide search for epigenetic variants that associate with complex phenotypes and have potential to identify novel genes and molecular pathways in common disease. However, unlike genetic variation, epigenetic variation can be dynamic, which has implications for EWAS methodology and design. This chapter discusses the analytical aspects of performing EWAS of DNA methylation changes in complex traits, as well as the potential to integrate genetic and epigenetic variation in the analysis of molecular mechanisms underlying complex phenotypes.

14.1 Introduction

Epigenome-wide association scans (EWAS) are large-scale studies of epigenetic changes in complex phenotypes. EWAS aim to systematically assess epigenetic variation throughout the genome, and test for association between epigenetic levels and complex traits. For example, a recent EWAS of rheumatoid arthritis characterized DNA methylation levels genome-wide and observed significantly lower DNA methylation levels at the major histocompatibility (MHC) region in rheumatoid arthritis cases, compared to controls (Liu et al. 2013). In many respects EWAS take a similar approach to genome-wide association studies (GWAS); however, they require further considerations in methodology, design, and

J.T. Bell (✉)

Department of Twin Research and Genetic Epidemiology, King's College, London, UK
e-mail: jordana.bell@kcl.ac.uk

interpretation due to the dynamic nature of epigenetic levels over time. At present, the majority of EWAS focus on DNA methylation, potentially because of the availability of genome-wide technologies to assay DNA methylation marks compared to assays available for other epigenetic mechanisms. With the advancement of molecular technologies EWAS will likely soon include additional epigenetic mechanisms, such as histone modifications and chromatin structure assays. Recently, EWAS of DNA methylation levels in complex disease have identified disease-associated differentially methylated regions (DMRs) for multiple traits (Bakulski et al. 2012; Bell et al. 2012; Breitling et al. 2011; Cheung et al. 2010; Gervin et al. 2012; Hasler et al. 2012; Javierre et al. 2010; Kaminsky et al. 2008; Mastroeni et al. 2010; Rakyan et al. 2010, 2011a; Teschendorff et al. 2010; Toperoff et al. 2012; Zhao et al. 2012), where modest changes in DNA methylation levels were associated with disease or quantitative trait levels. Furthermore, it has also been suggested that epigenetic variants may not only influence complex trait levels but may also contribute to phenotype variability (Feinberg and Irizarry 2010), both directly or through genetic–epigenetic interactions (Yang et al. 2012). This idea broadens the scope of potential EWAS methods towards integrative frameworks that can incorporate trait variance effects, thus moving beyond standard epigenetic-phenotype tests of association. The ultimate goal of EWAS is to obtain a mechanistic insight into the biological pathways involved in phenotype susceptibility and progression.

14.2 DNA Methylation

To date, the majority of epigenetic studies have focused on DNA methylation, which has unique features that may influence EWAS study design. DNA methylation is one of the most common and stable epigenetic mechanisms. DNA methylation can have downstream effects on gene expression and consequently play an important role in normal development and disease. Within the context of EWAS, a distinction should be made between the level of DNA methylation at the individual cell level, and the level of DNA methylation within a tissue sample from an individual subject, which consists of a population of cells (Fig. 14.1). It is typically DNA methylation levels within a sample from an individual that represent the sampling unit for the majority of EWAS approaches. DNA methylation levels and changes in DNA methylation levels can be heritable, both through mitosis and meiosis. However, methylation can also change over time and has been linked to environmental factors, such as smoking (Breitling et al. 2011). DNA methylation in mammals occurs on the cytosine base, predominantly but not exclusively in the context of cytosine-phosphate-guanine (CpG) dinucleotides (Ramsahoye et al. 2000). CpG dinucleotides tend to cluster in CpG islands (CGIs), which are regions of high CpG density. In general, DNA methylation of promoters is negatively associated with gene expression levels across the genome in cancer (*see* Jones and Baylin 2002) and healthy tissues (Eckhardt et al. 2006; Song et al. 2005),

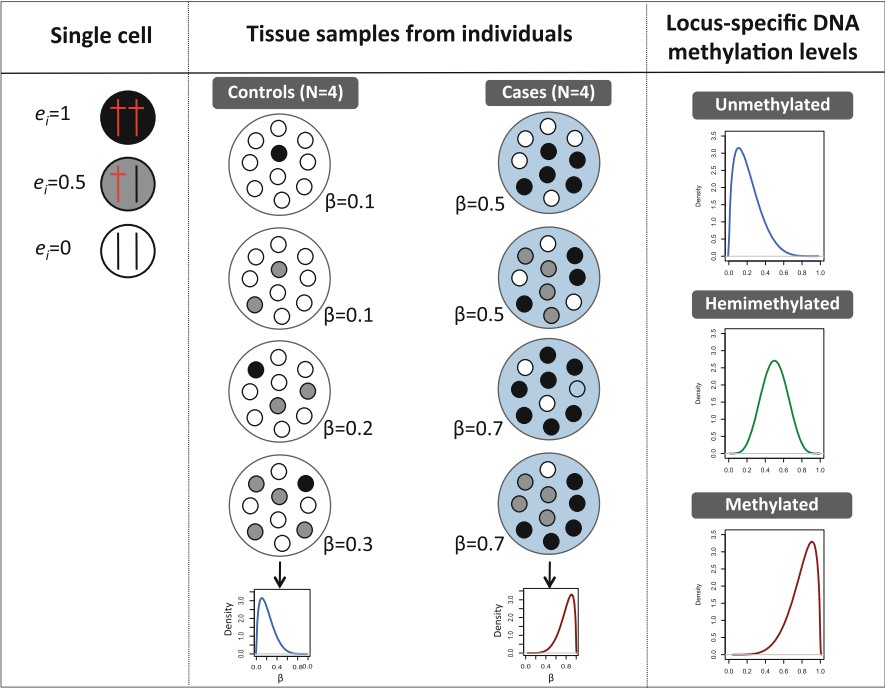


Fig. 14.1 An illustrative example of DNA methylation levels in the context of an epigenetic model of disease susceptibility. The *left panel* represents DNA methylation levels within a single cell. The *middle panel* represents tissue samples from subjects, where the overall sample DNA methylation level is expressed as β , or the allele frequency of the methylated allele in the population of cells within the tissue sample. In this hypothetical example the overall DNA methylation level in the tissue sample contributes to disease status in a case-control sample. The *right panel* represents the distribution of locus-specific DNA methylation levels in a sample of individuals for un-methylated, hemi-methylated, and methylated loci, which is representative of that observed on the Illumina InfiniumMethylation microarrays

while gene-body methylation has been positively correlated with expression (Aran et al. 2011; Jingo et al. 2012). DNA methylation is also known to be not only tissue specific but also cell type specific (Rakyan et al. 2008; Thompson et al. 2010; Heijmans et al. 2009; Ollikainen et al. 2010), which is one of the key features to consider in EWAS design choices. DNA methylation is strongly correlated with other epigenetic marks such as histone modification (Cedar and Bergman 2009) and other markers of open and closed chromatin structure (Dunham et al. 2012). These findings suggest that shared mechanisms of epigenetic regulation exist (Cedar and Bergman 2009; Fuks 2005; Zhou et al. 2010), and it has been proposed that transcriptionally silent chromatin may be a mark of de novo DNA methylation (Bird 2002), implying that at least in some cases DNA methylation may be the mark rather than trigger of chromatin-state change.

14.2.1 Genetic and Environmental Effects on DNA Methylation

DNA methylation at specific sites can be affected by genetic (Bell et al. 2012), environmental (Christensen et al. 2009; Gronniger et al. 2010), and stochastic variation (Bjornsson et al. 2008; Fraga et al. 2005). DNA methylation heritability has been shown in twin studies, where monozygotic (MZ) twins have more similar DNA methylation patterns than dizygotic (DZ) twins (Kaminsky et al. 2009), and from observations that methylation patterns segregate in families (Bjornsson et al. 2008). Recently, many genetic variants have been identified as methylation quantitative trait loci (me-QTLs), which are genetic polymorphisms that significantly associate with DNA methylation levels, predominantly in *cis* (Bell et al. 2011; Gibbs et al. 2010; Gamazon et al. 2012; Numata et al. 2012; Gertz et al. 2011; Schilling et al. 2009). DNA methylation levels have also been reported to be associated with environmental variants within individuals, such as smoking (Breitling et al. 2011). Environmental exposures in the parental generation, such as maternal smoking during pregnancy (Joubert et al. 2012) and parental diet (Heijmans et al. 2008; Vucetic et al. 2010), have also been associated with an individual's DNA methylation level at particular loci. Therefore, DNA methylation levels may change in response to environmental variation over time both within individuals and within families across generations (Wong et al. 2010b; Fraga et al. 2005; Bjornsson et al. 2008).

14.2.2 DNA Methylation Assays

Many approaches can be used to profile DNA methylation levels and a recent review by Laird (2010) gives an in-depth overview of their key features and potential sources of bias. Briefly, commonly used platforms in EWAS can be broadly divided into three groups (Rakyan et al. 2011b; Bell and Spector 2011; Laird 2010): (1) microarrays, (2) enrichment-based platforms, and (3) bisulfite-sequencing approaches. At present, the most commonly used microarray is the Illumina Infinium HumanMethylation450 (Illumina 450k) (Dedeurwaerder et al. 2011), which is the new version of the Illumina Infinium HumanMethylation27 (Illumina 27k) (Bibikova et al. 2009) and the Illumina GoldenGate Methylation Cancer Panel I (Illumina GoldenGate) arrays, and is based on genotyping of bisulfite-converted DNA. The assay profiles DNA methylation at ~485,000 CpG-sites predominantly located near genes, out of ~10⁷ possible CpG-sites across the genome. Unlike the Illumina 27k, the Illumina 450k does not target predominantly CpG-sites in CGIs, but also examines CGI shores, shelves, and regions of the genome of lower CpG density. In terms of potential sources of bias, as this approach is based on bisulfite-conversion, incomplete bisulfite-conversion and bisulfite-PCR can both be sources of bias, as well as the inability

to distinguish between different types of cytosine methylation (Huang et al. 2010), such as 5-methyl-cytosine (5-mC) and 5-hydroxy-methyl-cytosine (5-hmC). Despite incomplete genome coverage, the Illumina450k can be a cost-effective method to characterize DNA methylation patterns, and can allow standardized comparisons and meta-analysis across EWAS on a common platform (Rakyan et al. 2011b).

Enrichment-based platforms capture and sequence the methylated parts of the genome. Commonly used platforms include methylated DNA immunoprecipitation sequencing (MEDIP-seq), methylated DNA capture by affinity purification sequencing (MECAP-seq), and methylated DNA-binding domain sequencing (MBD-seq). MEDIP-seq uses an antibody specific for 5-mC to extract methylated fragments from sonicated DNA (Weber et al. 2005, 2007), while MECAP-seq (Brinkman et al. 2010) and MBD-seq (Weber et al. 2005; Zhang et al. 2006) use methyl-binding domain proteins to obtain methylated DNA. Affinity-based enrichment methods may exhibit a CpG density-dependent bias in binding affinity (Down et al. 2008) and a potential bias due to GC content. It is also difficult to obtain a measure of the overall level of DNA methylation in the sample. Importantly, these approaches do not provide single-CpG-level resolution of methylation, but are restricted by the size of DNA fragment (usually ~200 bp). On the other hand, the methods are rapid and efficient, provide good genome-wide coverage, and are able to measure allele-specific DNA methylation levels. Furthermore, specific enrichment-based approaches (for example, MeDIP-seq) can be designed to target specifically 5-mC or 5-hmC, and previous studies have used these approaches to examine non-CpG methylation levels (Bock et al. 2010; Li et al. 2010).

Bisulfite-sequencing methods include whole-genome bisulfite sequencing (WGBS) (Cokus et al. 2008; Lister and Ecker 2009) and reduced representation bisulfite sequencing (RRBS) (Gu et al. 2010). WGBS is currently the gold standard for assaying DNA methylation, because it can provide absolute levels of DNA methylation at single-CpG-site resolution at good genome-wide coverage, as well as an overall genome-wide level of DNA methylation for the sample, and it can also measure allele-specific methylation and DNA methylation levels at non-CpG-sites. However, this approach is also not ideal because certain regions of the genome are difficult to bisulfite sequence, either during the protocol or due to the alignment of bisulfite-converted reads. Potential sources of bias here include incomplete bisulfite-conversion and bisulfite-PCR bias, as well as the inability to distinguish between 5-mC and 5-hmC (Huang et al. 2010).

Sample requirements for DNA quality and quantity may impact the choice of assay (Laird 2010). For example, microarrays require less material (for example 500 ng for Illumina450k) (Bibikova et al. 2009; Dedeurwaerder et al. 2011) compared to enrichment-based or enzyme-based platforms (Laird 2010). The methylation profiling assays described here have been applied in multiple studies, but not all strategies have genome-wide coverage and different technical and biological factors (for example, CpG density) may bias the estimated DNA methylation levels. Therefore, it is important to validate DNA methylation levels at key regions of interest using multiple assays.

14.3 Design and Analysis of EWAS

EWAS study design and analysis require consideration due to the unique features of DNA methylation, which can be tissue specific, variable over time, and may both contribute to and result from phenotypic changes. Furthermore, assay sensitivity, coverage, and study design will also impact power to detect disease-associated DMRs (Fig. 14.2).

14.3.1 Tissue Specificity

DNA methylation variants can be tissue specific or shared across tissues (Gibbs et al. 2010; Gamazon et al. 2012; Numata et al. 2012). It is important to identify and sample the tissue that is most relevant for the trait. However, often the most appropriate tissue may not be available or easily accessible, and for many studies only whole-blood DNA will be available. The suitability of whole blood as a

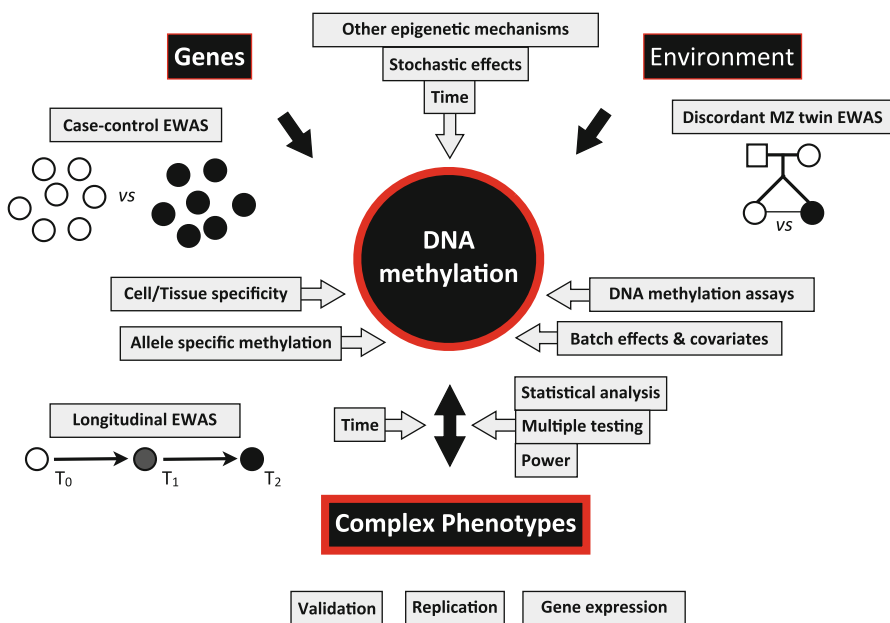


Fig. 14.2 Methodological considerations for EWAS study designs. The *top panel* lists the major factors that may influence DNA methylation levels. The *middle panel* represents a list of factors that may impact the measurement of DNA methylation levels. The *bottom panel* represents additional factors that may play a role in identifying and interpreting the observed associations between DNA methylation levels and disease. Three common types of EWAS study designs are incorporated within this example

surrogate for DNA methylation in other tissues is extensively debated (Heijmans and Mill 2012; Rakyan et al. 2011b; Tsai et al. 2012). At CpG-sites where DNA methylation is heritable or has been linked to genetic variants or me-QTLs, methylation is more stable over time and conserved across tissues. However, CpG-sites with me-QTLs constitute a small proportion of sites genome-wide. Due to the availability of whole-blood samples, whole-blood DNA methylation is often used for first-stage EWAS analyses, and subsequent follow-up aims to assess methylation tissue specificity at the disease DMRs in multiple tissues of relevance to disease. To address the concern of cell heterogeneity in whole-blood DNA methylation EWAS, where possible whole-blood DNA methylation levels should be corrected for blood cell subtype proportions if data are available (Adalsteinsson et al. 2012; Houseman et al. 2012). Additional EWAS designs with consideration of tissue specificity apply to cancer studies, where one approach would be to compare DNA methylation levels from healthy and affected tissue from the same individual, as well as include healthy tissue from unaffected individuals as another control sample.

14.3.2 Bias and Covariates in Large-Scale DNA Methylation Datasets

All DNA methylation assays may be affected by technical sources of noise and bias. For example, batch effects have been described in both microarray (Bell et al. 2011) and sequencing studies (Laird 2010). For Illumina DNA methylation microarrays, samples processed during the same bisulfite conversion step, samples assayed on the same Illumina chip, or samples located on the same position across chips may exhibit more similar profiles of DNA methylation, resulting in batch effects. Furthermore, known covariates such as age, cell subtype sample heterogeneity, and environmental factors such as smoking and diet will also impact DNA methylation levels at specific loci. EWAS should be designed to minimize these effects, that is, to exclude the possibility of these technical and biological covariates becoming confounders for the disease of interest. One standard approach towards this goal should be to randomly allocate samples across potential batches, and to perform analyses that adjust for measured and unmeasured covariates. Initial data checks should examine the distribution of methylation signals across potential batches. Correlation patterns in genome-wide DNA methylation levels should be assessed across the sample of individuals and within cases and controls separately, as well as across genome-wide loci and within autosomes and sex chromosomes separately. Several methods to assess heterogeneity in high-throughput technologies should be applied, for example, previous studies have used surrogate variable analysis in gene expression (Leek and Storey 2007), or principal component analysis and other approaches (see Leek et al. 2010). Data analysis post data normalization should be applied, but at the key regions of interest, analysis using unprocessed signals should also be performed (Laird 2010). Computational

approaches to control for confounders and technical artifacts can increase power to detect biological effects. For example, accounting for measured and unmeasured sources of heterogeneity in gene expression data, using surrogate variable analyses, increased the power to detect disease-associated gene expression profiles in large-scale samples (Leek and Storey 2007). The application of these methods to high-throughput DNA methylation datasets can reduce spurious signals and incorrect conclusions, and increase the reproducibility of the results.

14.3.3 EWAS Study Designs

Several EWAS study designs have been proposed (Rakyan et al. 2011b; Tsai et al. 2012). Similar to GWAS, cross-sectional population- and family-based EWAS designs can be adopted and many such cohorts are already available from GWAS efforts. In a population-based EWAS such as a case–control study, cases and controls are selected retrospectively and DNA methylation patterns are compared to detect disease-related DMRs. If the aim of EWAS is to identify extreme methylation changes, DNA methylation comparison of pooled cases and controls may also be pursued (Docherty et al. 2009). As in GWAS, sample heterogeneity and population stratification are also likely to impact population-based EWAS findings, and statistical measures to assess and control for these effects should be taken into account. For example, at a small proportion of CpG-sites across the genome DNA methylation levels are heritable and may be subject to population stratification effects. Aside from stratification, it is also possible that EWAS in cases and controls will detect DMRs at heritable methylation sites. Therefore, cross-sectional population-based EWAS may detect genetic associations with the phenotype that are mediated through DNA methylation.

In cross-sectional family-based designs, the disease-discordant identical twin design is often used in epigenetic studies of disease (Bell and Saffery 2012). MZ twins share nearly all genetic variants and many environmental and lifestyle factors, but have different epigenomes. Disease-discordant MZ twins have been used to assess the contribution of environmental, lifestyle, and epigenetic risk factors to many complex traits and phenotypes (see Bell and Spector 2011). The aim of twin-based EWAS is to clarify potential nongenetic epigenetic changes present in the case, but not control, twin. Other family-based study designs include comparisons across sib and half-sib pairs, and across parent–offspring pairs, trios, and multigenerational families, which may reveal the extent of genetic regulation of the epigenetic mark at a particular locus within and across generations. However, it is often difficult to obtain sufficient family-based samples for EWAS, particularly for rare diseases, for example, only 1 in 250 individuals has an identical twin (Bulmer 1970).

The main disadvantage of cross-sectional population- and family-based EWAS is that the timing of the methylation modification with respect to disease onset cannot be established, because methylation can be both causal and also a

consequence of disease. The only EWAS study design that can help address this in disease susceptibility is the longitudinal design, where DNA methylation levels at a locus of interest are assessed prior to and after the development of disease. These studies aim to identify DNA methylation changes that occur prior to disease onset and are therefore potentially causal to disease, and distinguish these from methylation variants that occur only after disease onset and are therefore likely consequences of disease progression. In addition, they allow for the assessment of the temporal variability in DNA methylation itself, which is also of biological interest. However, longitudinal samples are difficult to obtain. Furthermore, the optimal time of sampling with respect to disease onset is unknown, for example, is a 10-year window prior to and post disease onset the most optimal timescale? The answer is likely to depend on the age at disease onset, because there is evidence that rates of age-related changes in DNA methylation are higher in younger compared to older individuals (Talens et al. 2010; Tsai et al. 2012; Wong et al. 2010a). However, in practice nearly all longitudinal studies will be limited by the availability of longitudinal cohort samples.

Combining multiple types of EWAS study designs is perhaps the most useful approach to epigenetic studies of disease. For example, discovery-stage EWAS may start with a large case-control sample. If disease DMRs are identified, the second stage of analyses may address the potential causes of the DMRs, that is, whether the DMRs have a genetic or a nongenetic origin. This follow-up may include integrated genetic-epigenetic phenotype analysis in the original case-control sample, and disease-discordant twin analysis of the peak DMRs. The third stage would be longitudinal analyses to establish whether these DMRs are temporally stable, and potentially causal or consequences of disease (Fig. 14.2).

14.3.4 EWAS Power

Power to detect DMRs in EWAS depends on many factors including study design, sample size, DMR effect, interindividual variability in DNA methylation levels, methylation assay coverage and sensitivity, and stability of the DNA methylation variant over time. Relatively few studies have assessed EWAS power to date. In twins, EWAS published to date have used either small samples ($n < 5$) with high-resolution approaches such as bisulfite sequencing (Baranzini et al. 2010) or lower resolution assays such as Illumina450K or Illumina27K, with modest sample sizes ($n = 10-25$) (Bell et al. 2012; Dempster et al. 2010; Gervin et al. 2012; Hasler et al. 2012; Rakyan et al. 2011a). Several studies have estimated locus-specific power estimates for the disease-discordant twin design. Kaminsky et al. (2008) estimated power for a particular locus in a genome-wide methylation assay targeting CpG island regulatory elements. The results suggested that reasonable power to detect moderate effects could be obtained with a small sample of disease-discordant twins. For example, greater than 80 % power to detect a 1.15-fold change in the methylation signal was obtained with a sample size of 21 twin

pairs. However, the majority of currently used methylation assays assess single CpG-sites. Formal power calculations for more extensive genome-wide coverage at single CpG-site resolution have not yet been reported in twins. Preliminary estimates report low (35 %) to reasonable (>80 %) power to detect DMRs at specific CpG-sites at methylation differences of 5–6 % between affected and unaffected twins in 20–22 disease-discordant twin pairs (Bell et al. 2012; Dempster et al. 2010).

In case–control EWAS, two recent studies have explored power to detect DMRs. Wang (2011) compared the performance of a novel approach to standard parametric and nonparametric tests in EWAS, and suggested that greater power can be achieved if not only mean differences but also the variance of methylation levels are taken into account. The study observed that 86.6 % power was achieved to detect a DMR with a mean case–control difference of 9 % in methylation levels, using 250 cases and 250 controls. Rakyan et al. (2011b) also found that taking into account not only the difference but also the variance in methylation levels will impact power results. The authors estimated that 88 % power can be obtained to detect a DMR with methylation odds ratios (methOR) of 1.25 and sample size of 800 cases and 800 controls at a significance level of 10^{-6} if the locus is primarily methylated, but power is reduced (<10 %) with greater variability in methylation levels. The methOR is defined as the odds for a strand of DNA from a case to be methylated divided by the same odds for controls. Power estimates of 80 % could be achieved with a high methOR of 2.11 in at least 200 cases and 200 controls, but for a more realistic methOR of 1.49, increased sample sizes of at least 800 cases and 800 controls were required (Rakyan et al. 2011b).

14.3.5 EWAS Significance Thresholds

EWAS significance threshold levels should account for multiple testing. One approach is to consider the total number of CpG-sites examined. For example the Illumina 450k array includes ~485,000 CpG-sites across the genome, resulting in a Bonferroni-adjusted significance threshold of $P \sim 10^{-7}$. However, DNA methylation levels at nearby (within 1–2 kb) CpG-sites are correlated (co-methylation) (Bell et al. 2011; Eckhardt et al. 2006), which suggests that a Bonferroni correction is likely overconservative. Criteria for genome-wide significance thresholds in EWAS have not yet been established. Therefore, alternative methods to assess genome-wide significance should be considered, including false discovery rate (FDR) correction for multiple comparisons, or permutation-based approaches.

14.3.6 Validation

To date, most DNA methylation studies have used bisulfite sequencing for validation of selected DMRs. However, large-scale EWAS may require custom validation assays. A commonly used custom DNA methylation validation assay may be

bisulfite sequencing based, and would include the identified disease-related DMRs, genes near known GWAS signals for this disease, and potentially also other related disease DMRs, certain tissue-specific CpG-sites of tissue relevance to the disease of interest, as well as heritable CpG-sites, CpG-sites associated with environmental variants, and highly variable methylation regions. On the other hand, designing disease-specific validation assays may also be appropriate, for example, a cancer-specific panel may target CpG island shores, cancer-DMRs, and cancer-related hypo-methylation blocks (Hansen et al. 2011; Berman et al. 2012), while common-disease panels may include promoter-specific regions and previously identified disease-associated DMRs.

14.3.7 Replication

Replication guidelines for GWAS are clearly established and will be even more necessary in EWAS due to the dynamic nature of DNA methylation variation and the potential bias and noise in measuring DNA methylation levels genome-wide. Further complications that EWAS replication should address are DMR tissue specificity and underlying cause. For example, for a genetically driven DMR identified in a discovery case-control study, replication in disease-discordant twins may not be appropriate. Further details on replication considerations and guidelines for EWAS have been previously been discussed in more detail elsewhere (see Rakyan et al. 2011b).

14.4 After an EWAS Study: Further Investigations

Once EWAS DMRs have been identified, several follow-up analyses should be considered as routine steps. These include validation of the DNA methylation signal by a different assay, replication of the DMR effect in an independent sample, and longitudinal studies to assess the temporal stability and timing of the DMR relative to disease onset. After these initial analyses, further work can focus on understanding the role of the DMR in the trait.

14.4.1 Genetic–Epigenetic Analyses

Integrating genetic and epigenetic information in a shared analytical framework can help to assess whether the DMR mediates genotype–phenotype associations. In this case, DMR effects are likely to be detected in a sample of unrelated individuals, but may not be identified in samples of disease-discordant MZ twins. Examples of integrative analyses include genotype–epigenotype effects in rheumatoid arthritis

(Liu et al. 2013) and combining me-QTL, GWAS, and DMR findings in age-related phenotypes (Bell et al. 2012). Although genetic–epigenetic results imply causality, this is not necessarily the case, because it is possible that genetic associations lead to the phenotype, which in turn drives changes in methylation and alters gene expression as a consequence.

14.4.2 Causality

To date, longitudinal EWAS are the only design that can inform causality of DMR effects on the phenotype. If longitudinal samples are not available, case–control and twin EWAS follow-up in this area may include Mendelian randomization approaches (Relton and Davey Smith 2010), intervention trials, and causal inference analyses, for example, using structural equation modeling, Bayesian network analysis, or multivariate phenotype analysis methods. However, in all cases, to conclusively infer causal effects of DNA methylation on the trait, experimental evidence will be necessary. For some disease DMRs, it may be possible to manipulate DNA methylation levels at the DMR of interest in model organisms in the appropriate tissue and at the appropriate developmental stage, but for other traits such as psychiatric disorders experimental studies may prove challenging.

14.4.3 Gene Expression

Gene expression at the genes surrounding the DMR variant can help clarify function (Heijmans and Mill 2012). If expression data are not available for relevant tissues, comparing methylation and expression across multiple tissues may inform the tissue specificity of the variant. The presence of transcription factor-binding sites in the vicinity of the DMR variant may also prove informative. Comparison of multiple levels of epigenetic regulation in a wider genomic region of the DMR may also inform the role of the variant, as previously discussed (Heijmans and Mill 2012).

14.4.4 Allele-Specific Methylation

Recent studies have surveyed levels of allele-specific methylation (ASM), and show that ASM occurs not only at known imprinted loci but is also prevalent throughout the genome (Fang et al. 2012; Kerkel et al. 2008). However, the majority of EWAS to date assume that the overall level of DNA methylation at a locus of interest may influence the phenotype, yet more complex epigenetic risk models can also exist. For example, in allele-specific methylation, only one allele at

a heterozygous locus may be methylated, and this methylation could alter gene activity in *cis* and affect the phenotype of interest. Such effects may not be detected in standard EWAS using assays that measure overall levels of methylation at a locus, such as the Illumina450k array. Technologies that can detect allele-specific methylation (for example, MeDIP-seq and WGBS) will be useful to establish the extent of ASM effects on phenotypes.

14.4.5 EWAS for Phenotype Variability

It has been proposed that epigenetic variants could influence not only phenotype level but also phenotype variance (Feinberg and Irizarry 2010). Similar effects have been observed for genetic variants and several GWAS of phenotype variability have been performed across species (Hulse and Cai 2013; Ronnegard and Valdar 2011; Surakka et al. 2012; Yang et al. 2012). For example, recently a genetic variant in the *FTO* gene was identified to influence variability in body mass composition (BMI) in humans (Yang et al. 2012) and several mechanisms were proposed to explain allelic effects on BMI variance, including epigenetic effects.

EWAS study designs for trait variance can include comparisons of DNA methylation to phenotype variance in large-scale samples of unrelated individuals, and comparisons of MZ twin-pair discordance in DNA methylation and phenotype. Furthermore, it is also worth pursuing genotype–epigenotype integrative analyses in this setting because genetic effects on phenotype variability may be mediated by epigenetic variation (Feinberg and Irizarry 2010). It is also possible that a genetic variant that impacts phenotype variance interacts with an epigenetic variant, that is, DNA methylation may determine phenotype levels only in individuals of genotypes associated with greater phenotypic variance. This presents potential for novel types of analyses further integrating genetic and epigenetic data to understand the mechanisms involved in human complex traits.

14.5 Conclusion

EWAS provide promising methods to detect disease-associated epigenetic variation, but also highlight methodological challenges in studying dynamic epigenetic marks. These include the need to perform longitudinal EWAS, as well as validation and replication in large samples. Ultimately, integrative analysis across multiple sources of genomic, epigenomic, functional, and phenotypic data will help disentangle the biological mechanisms involved in human complex disease.

Acknowledgments I would like to acknowledge Pei-Chien Tsai's contribution to Fig. 14.1.

References

- Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, Gudnason V (2012) Heterogeneity in white blood cells has potential to confound DNA methylation measurements. *PLoS One* 7(10):e46705. doi:[10.1371/journal.pone.0046705](https://doi.org/10.1371/journal.pone.0046705)
- Aran D, Toperoff G, Rosenberg M, Hellman A (2011) Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet* 20(4):670–680, doi:[ddq513](https://doi.org/10.1093/hmg/ddq513)
- Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, Albin RL, Hu H, Rozek LS (2012) Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. *J Alzheimers Dis* 29(3):571–588. doi:[10.3233/JAD-2012-111223](https://doi.org/10.3233/JAD-2012-111223)
- Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo S, Kwok PY, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF (2010) Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464(7293):1351–1356, doi:[nature08990](https://doi.org/10.1038/nature08990)
- Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol* 12(1):R10, doi:[gb-2011-12-1-r10](https://doi.org/10.1186/gb-2011-12-1-r10)
- Bell JT, Saffery R (2012) The value of twins in epigenetic epidemiology. *Int J Epidemiol* 41(1):140–150. doi:[10.1093/ije/dyr179](https://doi.org/10.1093/ije/dyr179)
- Bell JT, Spector TD (2011) A twin approach to unraveling epigenetics. *Trends Genet* 27(3):116–125. doi:[10.1016/j.tig.2010.12.005](https://doi.org/10.1016/j.tig.2010.12.005)
- Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, Mangino M, Zhai G, Zhang F, Valdes A, Shin SY, Dempster EL, Murray RM, Grundberg E, Hedman AK, Nica A, Small KS, Dermitzakis ET, McCarthy MI, Mill J, Spector TD, Deloukas P (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* 8(4):e1002629. doi:[10.1371/journal.pgen.1002629](https://doi.org/10.1371/journal.pgen.1002629)
- Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, Van Den Berg D, Laird PW (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* 44(1):40–46. doi:[10.1038/ng.969](https://doi.org/10.1038/ng.969)
- Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL (2009) Genome-wide DNA methylation profiling using infinium(R) assay. *Epigenomics* 1(1):177–200. doi:[10.2217/epi.09.14](https://doi.org/10.2217/epi.09.14)
- Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16(1):6–21. doi:[10.1101/gad.947102](https://doi.org/10.1101/gad.947102)
- Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, Yu W, Rongione MA, Ekstrom TJ, Harris TB, Launer LJ, Eiriksdottir G, Leppert MF, Sapienza C, Gudnason V, Feinberg AP (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 299(24):2877–2883, doi:[299/24/2877](https://doi.org/10.1001/jama.299.24.2877)
- Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 28(10):1106–1114, doi:[nbt.1681](https://doi.org/10.1038/nbt.1681)
- Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 88(4):450–457, doi:[S0002-9297\(11\)00095-4](https://doi.org/10.1016/j.ajhg.2011.03.005)
- Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG (2010) Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* 52(3):232–236. doi:[10.1016/j.ymeth.2010.06.012](https://doi.org/10.1016/j.ymeth.2010.06.012)
- Bulmer MG (1970) *The biology of twinning in man*. Clarendon Press, Oxford

- Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10(5):295–304. doi:nrg2540
- Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM, Chan WY (2010) Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. *Br J Cancer* 102(2):419–427. doi:10.1038/sj.bjc.6605505
- Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R, Sugarbaker DJ, Yeh RF, Wiencke JK, Kelsey KT (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* 5(8):e1000602. doi:10.1371/journal.pgen.1000602
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452(7184):215–219. doi:10.1038/nature06745
- Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F (2011) Evaluation of the Infinium methylation 450K technology. *Epigenomics* 3(6):771–784. doi:10.2217/epi.11.105
- Dempster E, Pidsley R, Schalkwyk L, Touloupoulou T, Picchioni M, Kravariti E, Murray R, Mill J (2010) Methylomic profiling in twins discordant for major psychosis. *Twin Res Hum Genet* 13(3):253
- Docherty SJ, Davis OS, Haworth CM, Plomin R, Mill J (2009) Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics Chromatin* 2(1):3. doi:10.1186/1756-8935-2-3
- Down TA, Rakyen VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Backdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJ, Durbin R, Tavare S, Beck S (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol* 26(7):779–785. doi:nb1414
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Kheradpour P, Lassman T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SC, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LA, Adams LB, Kelly CJ, Zhang J, Wexler JR, Good PJ, Feingold EA, Crawford GE, Dekker J, Elinitzki L, Farnham PJ, Giddings MC, Gingeras TR, Guigo R, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Stamatoyannopoulos JA, Tennebaum SA, Weng Z, White KP, Wold B, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie L, Chen X, Mikkelsen TS, Gillespie S, Goren A, Ram O, Zhang X, Wang L, Issner R, Coyne MJ, Durham T, Ku M, Truong T, Eaton ML, Dobin A, Lassmann T, Tanzer A, Lagarde J, Lin W, Xue C, Williams BA, Zaleski C, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Batut P, Bell I, Bell K, Chakraborty S, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Li G, Luo OJ, Park E, Preall JB, Presaud K, Ribeca P, Robyr D, Ruan X, Sammeth M, Sandu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Hayashizaki Y, Reymond A, Antonarakis SE, Hannon GJ, Ruan Y, Carninci P, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP, Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kirkup VM, Meyer LR, Long JC, Maddren M, Raney BJ, Grasfeder LL, Giresi PG, Battenhouse A, Sheffield NC, Showers KA, London D, Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZ, Mieczkowski PA, Mieczkowska JO, Liu Z, McDaniel RM, Ni Y, Rashid NU, Kim MJ, Adar S, Zhang Z, Wang T, Winter D, Keefe D, Iyer VR, Sandhu KS, Zheng M, Wang P, Gertz J, Vielmetter J, Partridge EC, Varley KE, Gasper C, Bansal A,

- Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, Muratet MA, Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KI, Pusey B, DeSalvo G, Parker SL, Balasubramanian S, Davis NS, Meadows SK, Eggleston T, Newberry JS, Levy SE, Absher DM, Wong WH, Blow MJ, Visel A, Pennachio LA, Elnitski L, Petrykowska HM, Abyzov A, Aken B, Barrell D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Davidson C, Despacio-Reyes G, Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA, Hunt T, Jungreis I, Kay M, Khurana E, Leng J, Lin MF, Loveland J, Lu Z, Manthavadi D, Mariotti M, Mudge J, Mukherjee G, Notredame C, Pei B, Rodriguez JM, Saunders G, Shoner A, Searle S, Sisu C, Snow C, Steward C, Tapanari E, Tress ML, van Baren MJ, Washietl S, Wilming L, Zadissa A, Zhengdong Z, Brent M, Haussler D, Valencia A, Raymond A, Addleman N, Alexander RP, Auerbach RK, Bettinger K, Bhardwaj N, Boyle AP, Cao AR, Cayting P, Charos A, Cheng Y, Eastman C, Euskirchen G, Fleming JD, Grubert F, Habegger L, Hariharan M, Harmanci A, Iyenger S, Jin VX, Karczewski KJ, Kasowski M, Lacroute P, Lam H, Larnar-Vincent N, Lian J, Lindahl-Allen M, Min R, Miotto B, Monahan H, Moqtaderi Z, Mu XJ, O'Geen H, Ouyang Z, Patacsil D, Raha D, Ramirez L, Reed B, Shi M, Slifer T, Witt H, Wu L, Xu X, Yan KK, Yang X, Struhl K, Weissman SM, Tenebaum SA, Penalva LO, Karmakar S, Bhanvadia RR, Choudhury A, Domanus M, Ma L, Moran J, Victorsen A, Auer T, Centarin L, Eichenlaub M, Gruhl F, Heerman S, Hoeckendorf B, Inoue D, Kellner T, Kirchmaier S, Mueller C, Reinhardt R, Schertel L, Schneider S, Sinn R, Wittbrodt B, Wittbrodt J, Jain G, Balasundaram G, Bates DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen RS, Boatman L, Haugen E, Humbert R, Johnson AK, Johnson EM, Kuttyavin TM, Lee K, Lotakis D, Maurano MT, Neph SJ, Neri FV, Nguyen ED, Qu H, Reynolds AP, Roach V, Rynes E, Sanchez ME, Sandstrom RS, Shafer AO, Stergachis AB, Thomas S, Vernot B, Vierstra J, Vong S, Weaver MA, Yan Y, Zhang M, Akey JA, Bender M, Dorschner MO, Groudine M, MacCoss MJ, Navas P, Stamatoyannopoulos G, Stamatoyannopoulos JA, Beal K, Brazma A, Flicek P, Johnson N, Lusk M, Luscombe NM, Sobral D, Vaquerizas JM, Batzoglou S, Sidow A, Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Miller W, Bickel PJ, Banfai B, Boley NP, Huang H, Li JJ, Noble WS, Bilmes JA, Buske OJ, Sahu AO, Kharchenko PV, Park PJ, Baker D, Taylor J, Lochovsky L (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Little J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38(12):1378–1385
- Fang F, Hodges E, Molaro A, Dean M, Hannon GJ, Smith AD (2012) Genomic landscape of human allele-specific DNA methylation. *Proc Natl Acad Sci U S A* 109(19):7332–7337. doi:[10.1073/pnas.1201310109](https://doi.org/10.1073/pnas.1201310109)
- Feinberg AP, Irizarry RA (2010) Evolution in health and medicine Sackler colloquium: stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107(Suppl 1):1757–1764. doi:[10.1073/pnas.0906183107](https://doi.org/10.1073/pnas.0906183107)
- Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* 102(30):10604–10609. doi:[0500398102](https://doi.org/10.1073/pnas.0500398102)
- Fuks F (2005) DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* 15(5):490–495. doi:[S0959-437X\(05\)00132-2](https://doi.org/10.1016/j.cop.2005.08.001)
- Gamazon ER, Badner JA, Cheng L, Zhang C, Zhang D, Cox NJ, Gershon ES, Kelsoe JR, Greenwood TA, Nievergelt CM, Chen C, McKinney R, Shilling PD, Schork NJ, Smith EN, Bloss CS, Nurnberger JI, Edenberg HJ, Foroud T, Koller DL, Scheftner WA, Coryell W, Rice J, Lawson WB, Nwulia EA, Hipolito M, Byerley W, McMahon FJ, Schulze TG, Berrettini

- WH, Potash JB, Zandi PP, Mahon PB, McInnis MG, Zollner S, Zhang P, Craig DW, Szelinger S, Barrett TB, Liu C (2012) Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry*. doi:[10.1038/mp.2011.174](https://doi.org/10.1038/mp.2011.174)
- Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, Parker SL, Kucera KS, Willard HF, Myers RM (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet* 7(8):e1002228. doi:[10.1371/journal.pgen.1002228](https://doi.org/10.1371/journal.pgen.1002228)
- Gervin K, Vigeland MD, Mattingsdal M, Hammero M, Nygard H, Olsen AO, Brandt I, Harris JR, Undlien DE, Lyle R (2012) DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. *PLoS Genet* 8(1):e1002454. doi:[10.1371/journal.pgen.1002454](https://doi.org/10.1371/journal.pgen.1002454)
- Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 6(5):e1000952. doi:[10.1371/journal.pgen.1000952](https://doi.org/10.1371/journal.pgen.1000952)
- Gronniger E, Weber B, Heil O, Peters N, Stab F, Wenck H, Korn B, Winnefeld M, Lyko F (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet* 6(5):e1000971. doi:[10.1371/journal.pgen.1000971](https://doi.org/10.1371/journal.pgen.1000971)
- Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, Tomazou E, Gnirke A, Lander ES, Meissner A (2010) Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nat Methods* 7(2):133–136. doi:[nmeth.1414](https://doi.org/10.1038/nmeth.1414)
- Hansen KD, Timp W, Bravo HC, Sabuncian S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43(8):768–775. doi:[10.1038/ng.865](https://doi.org/10.1038/ng.865)
- Hasler R, Feng Z, Backdahl L, Spehlmann ME, Franke A, Teschendorff A, Rakyan VK, Down TA, Wilson GA, Feber A, Beck S, Schreiber S, Rosenstiel P (2012) A functional methylome map of ulcerative colitis. *Genome Res* 22(11):2130–2137. doi:[10.1101/gr.138347.112](https://doi.org/10.1101/gr.138347.112)
- Heijmans BT, Mill J (2012) Commentary: the seven plagues of epigenetic epidemiology. *Int J Epidemiol* 41(1):74–78. doi:[10.1093/ije/dyr225](https://doi.org/10.1093/ije/dyr225)
- Heijmans BT, Tobi EW, Lumey LH, Slagboom PE (2009) The epigenome: archive of the prenatal environment. *Epigenetics* 4(8):526–531
- Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* 105(44):17046–17049. doi:[10.1073/pnas.0806560105](https://doi.org/10.1073/pnas.0806560105)
- Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86. doi:[10.1186/1471-2105-13-86](https://doi.org/10.1186/1471-2105-13-86)
- Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 5(1):e8888. doi:[10.1371/journal.pone.0008888](https://doi.org/10.1371/journal.pone.0008888)
- Hulse AM, Cai JJ (2013) Genetic variants contribute to gene expression variability in humans. *Genetics* 193(1):95–108. doi:[10.1534/genetics.112.146779](https://doi.org/10.1534/genetics.112.146779)
- Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JJ, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, Jacinto FV, Lopez-Longo FJ, Dopazo J, Forn M, Peinado MA, Carreno L, Sawalha AH, Harley JB, Siebert R, Esteller M, Miller FW, Ballestar E (2010) Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. *Genome Res* 20(2):170–179. doi:[gr.100289.109](https://doi.org/10.1101/002891)
- Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK (2012) On the presence and role of human gene-body DNA methylation. *Oncotarget* 3(4):462–474. doi:[497](https://doi.org/10.1038/nrg816)
- Jones PA, Baylin SB (2002) The fundamental role of epigenetic events in cancer. *Nat Rev Genet* 3(6):415–428. doi:[10.1038/nrg816](https://doi.org/10.1038/nrg816)

- Joubert BR, Haberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, Huang Z, Hoyo C, Midttun O, Cupul-Uicab LA, Ueland PM, Wu MC, Nystad W, Bell DA, Peddada SD, London SJ (2012) 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120(10):1425–1431. doi:[10.1289/ehp.1205412](https://doi.org/10.1289/ehp.1205412)
- Kaminsky Z, Petronis A, Wang SC, Levine B, Ghaffar O, Floden D, Feinstein A (2008) Epigenetics of personality traits: an illustrative study of identical twins discordant for risk-taking behavior. *Twin Res Hum Genet* 11(1):1–11. doi:[10.1375/twin.11.1.1](https://doi.org/10.1375/twin.11.1.1)
- Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C, McRae AF, Visscher PM, Montgomery GW, Gottesman II, Martin NG, Petronis A (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* 41(2):240–245. doi:[ng.286](https://doi.org/10.1038/ng.286) [pii]
- Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 40(7):904–908
- Laird PW (2010) Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11(3):191–203. doi:[10.1038/nrg2732](https://doi.org/10.1038/nrg2732)
- Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 11(10):733–739. doi:[10.1038/nrg2825](https://doi.org/10.1038/nrg2825)
- Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 3(9):1724–1735. doi:[10.1371/journal.pgen.0030161](https://doi.org/10.1371/journal.pgen.0030161)
- Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, Han X, Chen Q, Zhang X, Wang J (2010) Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods* 52(3):203–212. doi:[10.1016/j.ymeth.2010.04.009](https://doi.org/10.1016/j.ymeth.2010.04.009)
- Lister R, Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 19(6):959–966
- Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP (2013) Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* 31(2):142–147. doi:[10.1038/nbt.2487](https://doi.org/10.1038/nbt.2487)
- Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J (2010) Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol Aging* 31(12):2025–2037. doi:[10.1016/j.neurobiolaging.2008.12.005](https://doi.org/10.1016/j.neurobiolaging.2008.12.005)
- Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, Colantuoni C, Weinberger DR, Kleinman JE, Lipska BK (2012) DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet* 90(2):260–272. doi:[10.1016/j.ajhg.2011.12.020](https://doi.org/10.1016/j.ajhg.2011.12.020)
- Ollikainen M, Smith KR, Joo EJ, Ng HK, Andronikos R, Novakovic B, Abdul Aziz NK, Carlin JB, Morley R, Saffery R, Craig JM (2010) DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Hum Mol Genet* 19(21):4176–4188. doi:[10.1093/hmg/ddq336](https://doi.org/10.1093/hmg/ddq336)
- Rakyan VK, Beyan H, Down TA, Hawa MI, Maslau S, Aden D, Daunay A, Busato F, Mein CA, Manfras B, Dias KR, Bell CG, Tost J, Boehm BO, Beck S, Leslie RD (2011a) Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet* 7(9):e1002300. doi:[10.1371/journal.pgen.1002300](https://doi.org/10.1371/journal.pgen.1002300)
- Rakyan VK, Down TA, Balding DJ, Beck S (2011b) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12(8):529–541. doi:[10.1038/nrg3000](https://doi.org/10.1038/nrg3000)
- Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM, Leslie RD, Deloukas P, Spector TD (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* 20(4):434–439. doi:[gr.103101.109](https://doi.org/10.1093/gr.103101.109) [pii]

- Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJ, Carter NP, Tavare S, Beck S (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res* 18(9):1518–1529. doi:[10.1101/gr.077479.108](https://doi.org/10.1101/gr.077479.108)
- Ramsahoye BH, Biniszkiwicz D, Lyko F, Clark V, Bird AP, Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci U S A* 97(10):5237–5242, doi:[10.1073/pnas.0408436102](https://doi.org/10.1073/pnas.0408436102)
- Relton CL, Davey Smith G (2010) Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med* 7(10):e1000356. doi:[10.1371/journal.pmed.1000356](https://doi.org/10.1371/journal.pmed.1000356)
- Ronnegard L, Valdar W (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* 188(2):435–447. doi:[10.1534/genetics.111.127068](https://doi.org/10.1534/genetics.111.127068)
- Schilling E, El Chartouni C, Rehli M (2009) Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res* 19(11):2028–2035. doi:[10.1101/gr.095562.109](https://doi.org/10.1101/gr.095562.109)
- Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, Held WA (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc Natl Acad Sci U S A* 102(9):3336–3341. doi:[10.1073/pnas.0408436102](https://doi.org/10.1073/pnas.0408436102)
- Surakka I, Whitfield JB, Perola M, Visscher PM, Montgomery GW, Falchi M, Willemsen G, de Geus EJ, Magnusson PK, Christensen K, Sorensen TI, Pietilainen KH, Rantanen T, Silander K, Widen E, Muiilu J, Rahman I, Liljedahl U, Syvanen AC, Palotie A, Kaprio J, Kyvik KO, Pedersen NL, Boomsma DI, Spector T, Martin NG, Ripatti S, Peltonen L (2012) A genome-wide association study of monozygotic twin-pairs suggests a locus related to variability of serum high-density lipoprotein cholesterol. *Twin Res Hum Genet* 1–9. doi:[S11832-012-0006-3](https://doi.org/10.1007/s11832-012-0006-3)
- Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, Putter H, Slagboom PE, Heijmans BT (2010) Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J* 24(9):3135–3144, doi:[fj.09-150490](https://doi.org/10.1096/fj.09-150490)
- Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, Campan M, Noushmehr H, Bell CG, Maxwell AP, Savage DA, Mueller-Holzner E, Marth C, Kocjan G, Gayther SA, Jones A, Beck S, Wagner W, Laird PW, Jacobs IJ, Widschwendter M (2010) Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res* 20(4):440–446, doi:[gr.103606.109 \[pii\]](https://doi.org/10.1101/j.1474-9726.2010.00577.x)
- Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM, Barzilai N (2010) Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell* 9(4):506–518. doi:[10.1111/j.1474-9726.2010.00577.x](https://doi.org/10.1111/j.1474-9726.2010.00577.x)
- Toperoff G, Aran D, Kark JD, Rosenberg M, Dubnikov T, Nissan B, Wainstein J, Friedlander Y, Levy-Lahad E, Glaser B, Hellman A (2012) Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. *Hum Mol Genet* 21(2):371–383. doi:[10.1093/hmg/ddr472](https://doi.org/10.1093/hmg/ddr472)
- Tsai PC, Bell JT, Spector TD (2012) Using epigenome-wide association scans of DNA methylation in age-related complex human traits. *Epigenomics* 4(5):511–526
- Vucetic Z, Kimmel J, Totoki K, Hollenbeck E, Reyes TM (2010) Maternal high-fat diet alters methylation and gene expression of dopamine and opioid-related genes. *Endocrinology* 151(10):4756–4764. doi:[10.1210/en.2010-0505](https://doi.org/10.1210/en.2010-0505)
- Wang S (2011) Method to detect differentially methylated loci with case-control designs using illumina arrays. *Genet Epidemiol* 35(7):686–694. doi:[10.1002/gepi.20619](https://doi.org/10.1002/gepi.20619)
- Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37(8):853–862, doi:[ng1598](https://doi.org/10.1038/ng1598)
- Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39(4):457–466. doi:[10.1038/ng1990](https://doi.org/10.1038/ng1990)

- Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J (2010a) A longitudinal study of epigenetic variation in twins. *Epigenetics* 5(6). doi:12226
- Wong CC, Caspi A, Williams B, Craig IW, Houts R, Ambler A, Moffitt TE, Mill J (2010b) A longitudinal study of epigenetic variation in twins. *Epigenetics* 5(6):516–526. doi:12226
- Yang J, Loos RJ, Powell JE, Medland SE, Speliotes EK, Chasman DI, Rose LM, Thorleifsson G, Steinthorsdottir V, Magi R, Waite L, Smith AV, Yerges-Armstrong LM, Monda KL, Hadley D, Mahajan A, Li G, Kapur K, Vitart V, Huffman JE, Wang SR, Palmer C, Esko T, Fischer K, Zhao JH, Demirkan A, Isaacs A, Feitosa MF, Luan J, Heard-Costa NL, White C, Jackson AU, Preuss M, Ziegler A, Eriksson J, Kutalik Z, Frau F, Nolte IM, Van Vliet-Ostaptchouk JV, Hottenga JJ, Jacobs KB, Verweij N, Goel A, Medina-Gomez C, Estrada K, Bragg-Gresham JL, Sanna S, Sidore C, Tyrer J, Teumer A, Prokopenko I, Mangino M, Lindgren CM, Assimes TL, Shuldiner AR, Hui J, Beilby JP, McArdle WL, Hall P, Haritunians T, Zgaga L, Kolcic I, Polasek O, Zemunik T, Oostra BA, Junttila MJ, Gronberg H, Schreiber S, Peters A, Hicks AA, Stephens J, Foad NS, Laitinen J, Pouta A, Kaakinen M, Willesmsen G, Vink JM, Wild SH, Navis G, Asselbergs FW, Homuth G, John U, Iribarren C, Harris T, Launer L, Gudnason V, O'Connell JR, Boerwinkle E, Cadby G, Palmer LJ, James AL, Musk AW, Ingelsson E, Psaty BM, Beckmann JS, Waeber G, Vollenweider P, Hayward C, Wright AF, Rudan I, Groop LC, Metspalu A, Khaw KT, van Duijn CM, Borecki IB, Province MA, Wareham NJ, Tardif JC, Huikuri HV, Cupples LA, Atwood LD, Fox CS, Boehnke M, Collins FS, Mohlke KL, Erdmann J, Schunkert H, Hengstenberg C, Stark K, Lorentzon M, Ohlsson C, Cusi D, Staessen JA, Van der Klauw MM, Pramstaller PP, Kathiresan S, Jolley JD, Ripatti S, Jarvelin MR, de Geus EJ, Boomsma DI, Penninx B, Wilson JF, Campbell H, Chanock SJ, van der Harst P, Hamsten A, Watkins H, Hofman A, Witteman JC, Zillikens MC, Uitterlinden AG, Rivadeneira F, Kiemeny LA, Vermeulen SH, Abecasis GR, Schlessinger D, Schipf S, Stumvoll M, Tonjes A, Spector TD, North KE, Lettre G, McCarthy MI, Berndt SI, Heath AC, Madden PA, Nyholt DR, Montgomery GW, Martin NG, McKnight B, Strachan DP, Hill WG, Snieder H, Ridker PM, Thorsteinsdottir U, Stefansson K, Frayling TM, Hirschhorn JN, Goddard ME, Visscher PM (2012) FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490(7419):267–272. doi:[10.1038/nature11401](https://doi.org/10.1038/nature11401)
- Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126(6):1189–1201. doi:[10.1016/j.cell.2006.08.003](https://doi.org/10.1016/j.cell.2006.08.003)
- Zhao J, Goldberg J, Bremner JD, Vaccarino V (2012) Global DNA methylation is associated with insulin resistance: a monozygotic twin study. *Diabetes* 61(2):542–546. doi:[10.2337/db11-1048](https://doi.org/10.2337/db11-1048)
- Zhou VW, Goren A, Bernstein BE (2010) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12(1):7–18. doi:nrg2905

Index

A

Agouti, 111, 112
 Allele-specific epigenetic modifications, 214
 Allele-specific expression, 49, 131–133, 137, 138, 214, 296
 Allele-specific methylation, 214, 290, 293, 296, 297, 323, 330–331
 Association analysis, 274, 279–282

C

Callipyge locus, 89–102
 Callipyge phenotype, 89–91, 99–100, 102
 Cancer epigenetics, 209–223
 Case-control, 155, 250, 254, 260, 266, 281, 282, 288, 298, 321, 326–330
 CGI. *See* CpG island (CGI)
 Childhood abuse, 183, 185, 194
 Childhood neglect, 183
 Chromatin looping, 43, 154, 210, 212
 Chromatin modifications, 46, 135, 186, 190, 192, 214, 222, 249
 Chromatin remodeling, 41, 119, 150, 151, 154, 210
 Chromosomal rearrangements, 147–149, 153, 155, 156, 159–160, 162, 163, 217
 Clonal heterogeneity, 139
 Clonality, 73, 74, 135–140
 CLPG locus, 91
 CNV. *See* Copy number variation (CNV)
 Complex traits, 123, 136, 319–331
 Copy number variation (CNV), 148–166, 214–215
 CpG dinucleotide, 17, 149, 217, 292–295, 320
 CpG island (CGI), 17–21, 23, 40, 44, 49, 52, 76, 78, 94, 110, 182, 186, 211, 292–297, 304, 305, 320, 322, 327, 329

D

Data integration, 300–303
 Demethylation, 10–16, 23, 26, 37, 44, 46, 108–111, 114, 186, 187
 Differentially methylated region (DMR), 17, 20, 21, 25, 26, 36–37, 92, 94, 120, 293, 297, 302, 304–306, 320–330
 DNA methylation, 4–26, 36, 38–40, 44, 45, 49, 52, 54, 72, 76, 78–80, 98, 99, 108–117, 119–123, 135, 149, 153, 154, 159, 165, 180–189, 191–195, 197, 211, 212, 221, 222, 258, 259, 290–297, 299–302, 306, 307, 319–331
 DNA methylome, 292–296, 304
 Dosage compensation, 75–81, 153

E

Early-life adversity, 179–187, 192, 194, 199
 Embryo lethality, 266–271, 273
 Endogenous retroviruses, 41, 110, 111, 114
 Epigenetic(s), 35–55, 107–123, 247–251, 253, 258–260
 heterogeneity, 136, 139
 inheritance, 25, 26, 110–115, 117–120, 122, 221, 235–237, 239–241, 243
 marks, 4, 36, 43, 52, 55, 78, 108–110, 113–117, 119–121, 123, 185, 210, 211, 221, 234, 236, 237, 243, 298, 321, 326, 331
 mechanisms, 35, 39, 54, 132, 133, 148–154, 162, 165, 180–182, 186, 188, 189, 191, 192, 290, 320
 modifications, 25, 44, 123, 184, 210, 214, 221, 234–238, 242, 243, 249, 250, 288–292, 295, 305
 patterning, 212–219, 222

Epigenetic(s) (*cont.*)

regulation, 148, 212–216, 218, 221
 reprogramming, 3–26, 36–39, 45, 46
 silencing, 91, 213, 214, 220, 221
 transmission, 4, 25, 237–240, 242, 243
 variation, 74, 112, 118, 119, 123, 233–243,
 319, 331

Epigenome(s), 4, 48, 55, 120, 123, 179–199,
 210, 214, 217, 223, 249, 287–307,
 319–331

Epigenome-wide association scans (EWAS),
 287–307, 319–331

Epigenomics, 11, 54, 55, 297, 299, 300–303,
 305–307, 331

Epimutations, 25, 121, 123, 210, 213, 215, 236,
 240, 242–243

Evolution, 114, 164, 165, 233–243, 282, 305

EWAS. *See* Epigenome-wide association
 studies (EWAS)

F

Familial aggregation, 234, 235, 240

Family based association, 267

Reproductive fitness, 75

G

Gametogenesis, 4, 24, 25, 108–113, 132, 211,
 236, 239

Gene-gene interactions, 249

Genetic epidemiology, 234, 247, 248

Genomic imprinting, 4, 11, 25, 114, 115,
 267, 292

H

Haplotype-specific methylation (HSM), 297,
 300, 301

Heritability, 68, 233–243, 253, 258, 259, 278,
 288, 300, 322

Heterochromatin, 9, 38, 42, 50, 81, 116, 153,
 181, 211, 303

Histone modifications, 7, 8, 10, 12, 23, 24,
 26, 37, 38, 40, 41, 43, 50, 55, 113,
 122, 152, 161, 180, 181, 188, 189,
 191, 192, 194, 210, 214, 241, 249,
 291, 320, 321

HPA axis, 182–184, 186, 192, 193

HSM. *See* Haplotype-specific methylation
 (HSM)

I

ICR. *See* Imprinting control region (ICR)

Imprint(ing), 91–95, 101, 266, 267, 271–273
 errors, 271–273

resetting, 11, 114, 115, 267, 268,
 271–274

Imprinting control region (ICR), 4, 11, 17, 37,
 94, 153, 160–162, 218

Incomplete penetrance, 149

L

Linkage analysis, 270–271, 273–275

Log-linear models, 255

M

Madumnal, 91–95, 97–99, 102

Maternal behavior, 183, 184, 192

Maternal effects, 116, 117, 248, 249,
 261, 270

me-QTL. *See* Methylation quantitative trait
 loci (me-QTL)

Methylation erasure, 9–12, 15, 25, 26

Methylation establishment, 18–23

Methylation profiling, 54, 217, 307, 323

Methylation quantitative trait loci (me-QTL),
 322, 325, 330

Methylome, 292–296, 298, 300, 302–305

Microdeletion, 155–158, 161

microRNAs (miRNAs), 24, 71, 110, 181, 210,
 212, 213, 215–216, 218, 219, 221

Monoallelic expression, 94, 131–140

Mosaicism, 64–75, 80, 95, 133, 139, 159

Muscular hypertrophy, 89, 90, 99, 100, 102

N

Neurodevelopmental disorders, 140,
 147, 166

P

Padumnal, 92, 94, 95, 97–99

Parental imprinting, 91, 93, 153, 159

Parent of origin effects, 101, 137, 160, 219,
 248, 249, 252, 255, 269

Pluripotency, 4, 7, 9, 12, 39–47

Polar overdominance, 89, 102

Primordial germ cell development, 3, 6, 13

Pseudo-imprinting, 101

S

Shared environment, 237, 239, 240, 243, 249, 251–253
Skewed X-chromosome inactivation, 64, –68, 70–75, 77
Somatic mutations, 75, 81, 139, 210, 213, 216, 218, 222, 290, 307
Suicide, 179, 180, 182, 185, 188–196

T

Tissue-specific epigenetic regulation, 36
Tissue-specific transcription, 35
Transcription factors, 9, 35, 39, 41, 42, 45, 47–54, 137, 156, 158, 184, 186, 189, 192, 195, 293, 297, 306, 330
Transgene(s), 110, 114, 116, 119
Transgenerational effects, 121, 221, 248–261
Transgenerational epigenetic inheritance, 26, 110–115, 117, 120, 122, 221, 237, 243
Transgenerational inheritance, 25, 119, 239

Translocation, 7, 14, 15, 23, 36, 65, 117, 119–121, 147, 148, 156, 159, 212, 217, 218
Transmission distortion, 114, 116
Transmission ratio distortion, 114, 116, 265–282
Transposable elements, 10, 12, 13, 17, 21–23, 25, 292
Tumorigenesis, 210, 213, 217, 220, 222

V

Variable expressivity, 149, 157

X

X-chromosome inactivation, 63, 81, 151, 153
X-inactivation escape, 75, 81
X-linked disease, 71, 74, 75, 79–81
X-linked mutation, 64, 71, 72